Entrapping hidden changes in nature

understanding the past with the power of recurrences

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

d o c t o r r e r u m n a t u r a l i u m
(Dr. rer. nat.)
im Fach Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
MSc.-Phys. Deniz Eroglu

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Elmar Kulke

Gutachter:

1. Prof. Dr. Dr. h.c. mult. Jürgen Kurths
2. Prof. Dr. Muhittin Mungan
3. Prof. Dr. Ricardo Luiz Viana

Tag der mündlichen Prüfung: 29.01.2016

to Funda and Ahmet Eroglu,
for the unconditional and unlimited love, support and belief...

Abstract

The study of climate change is a very important field of science, since life is directly affected by these changes. Many proxies, i.e. data previously preserved within speleothems, sediments, tree rings, corals, etc., have been observed and many theoretical models have been developed in order to gain a comprehensive understanding of past and future climate. Investigating the past climate changes leads to the forecasts of future possibilities. Therefore, enlightening the past is a keystone to answer questions about climate as well as vegetation, life, evolution, and nature. Paleoclimatology, the understanding of changes in climate during the history of Earth, has become a popular area of research at the intersection of climatology, chemistry, physics and mathematics. In order to extract the proxy information, many observations and mathematical studies have been performed.

In this thesis, I focus on detecting dynamical regime transitions in nonlinear dynamical systems as well as in climate proxies where they mark previous critical climate changes. The Poincaré recurrence theorem, a fundamental theory in nonlinear dynamics proves that certain systems will revisit a state which is infinitesimally close to a previous state in finite time, is the backbone of the work. The thesis begins with a brief methodological background and brings the recurrence plot (RP) to forefront as the main tool for the further analyses. The thesis consists of three main studies: (i) The formation of RPs naturally depends on a free parameter in the analysis given by the distance threshold. I propose an alternative definition by using a weighted variant of the RP, called weighted recurrence plot (wRP), which removes dependence on this free parameter. Moreover, the entropy of RP for some cases leads to inconsistent results, and a new entropy definition of wRP overcome this problem. (ii) Furthermore, I suggest a novel way to select the threshold for a recurrence network (RN) for a specific time series. Selecting the optimization parameters for a specific time series is very important for the performance of the analysis. (iii) In the following theoretical work, I introduce a new preprocessing technique to deal with the heterogeneousness of time series, since the RP is not directly applicable on such data sets and the proxies from speleothems, in general, are irregularly sampled.

These theoretical approaches are performed on prototypical models in order to test their feasibility. The synthetic time series are created in regard of the possible difficulties which can arise from real-world applications. The success of the methods is represented and discussed in detail. Based upon the results of pragmatic applications, I have applied the methods on real data sets. The dynamical regime transitions are detected on an electrochemical experiment and several paleoclimate speleothem proxies. The most striking application of this thesis, the behaviour of the East Asian and Indonesian–Australian summer monsoons throughout the Holocene has been the focus, although no explanation of millennial-scale relationships between the two has yet been offered. An understanding of any such phasing relationships would provide insight into the long-term variability of the broader East Asian–Indonesian–Australian monsoon regime. Applying the new techniques to Holocene speleothem proxy records from northwestern Australia and southern China, I demonstrate the existence of a bipolar seesaw relationship between the two regional monsoon systems and

show that this anti-phasing can be related to shifts in the ITCZ and related tropical precipitation in response to solar variability.

Among these presented approaches are suitable methods to investigate the dynamical transition and they can be used for different purposes in a large variety of time series analyses. For instance, these techniques can be used in several different disciplines which have heterogeneousness in their sources. Finally, after the relationship between the two proxies from northwestern Australia and southern China has been uncovered, in the future, it should be achievable to extend the study to create a large paleoclimate relationship network for the entire Earth by using the methods given in this Thesis.

Zusammenfassung

Das Studium des derzeitigen Klimawandels ist von hoher Relevanz da viele Menschen dessen direkten Folgen ausgesetzt sind. Dazu werden in der Regel sogenannte Proxydaten ausgewertet, d.h., in Speläothemen, Sedimenten, Baumringen oder auch Korallen enthaltene Informationen über vergangene Klimaverhältnisse. Diese Informationen gehen in die Modellentwicklung ein, um ein umfangreiches Verständnis des vergangenen als auch zukünftigen Klimas zu erlangen. Die Untersuchung vergangener Klimaänderungen führt zu Vorhersagen über mögliche zukünftige Klimaentwicklungen. Daher ist das Verständnis der Vergangenheit der Schlüssel zu Fragen zum Klima als auch zur Vegetation, dem Leben, der Evolution und der Natur allgemein. Die Paläoklimatologie, die solche Klimaveränderungen in der Erdgeschichte untersucht, hat sich zu einem sehr aktiven Forschungsfeld an der Schnittstelle von Klimatologie, Chemie, Physik und Mathematik entwickelt.

Diese Dissertation beschäftigt sich damit Übergänge zwischen dynamischen Regimen in nichtlinearen Systemen zu detektieren (alias kritischen Klimaveränderungen in der Erdgeschichte). Der Poincaré'sche Rekurrenzsatz ist hierbei der wichtigste Baustein. Dieser fundamentale Satz der nichtlinearen Dynamik besagt, dass ein Zustand solcher Systeme in endlicher Zeit seiner Anfangsbedingung wieder nahe kommt. Den Anfang macht eine kurze Einführung in die Methodik mit der Einführung der Rekurrenzplots (RP) als wichtigem Werkzeug für die folgende Analyse. Die genaue Gestalt eines RP hängt dabei von einem freien Parameter ab, dem Distanzschwellwert. Hier schlage ich eine neue Variante vor, den gewichteten Rekurrenzplot (wRP), welcher ohne Abhängigkeit von diesem Parameter auskommt. Weiterhin führt die klassische Entropie von RP in manchen Fällen zu inkonsistenten Ergebnissen. Eine neue Entropiedefinition im Falle der wRP löst dieses Problem. Darüberhinaus schlage ich eine neue Möglichkeit vor den Schwellwert für Rekurrenznetzwerke (RN) für eine gegebene Zeitreihe auszuwählen. Einen optimalen Schwellwert auszuwählen ist ausschlaggebend für das Ergebnis der Zeitreihenanalyse. In den folgenden theoretischen Ausführungen beschreibe ich daher eine neuartige Vorbehandlung von heterogenen (insbesondere irregulär gesampelten) Zeitreihen. Diese treten oft in Proxydaten, etwa Speläothemen, auf und lassen sich nicht direkt mit RP analysieren. Die Anwendbarkeit dieser theoretischen Verbesserungsansätze wird an prototypischen Modellen getestet. Dabei werden synthetische Zeitreihen generiert, welche die in der Realität auftretenden Schwierigkeiten abbilden sollen. Der Erfolg der einzelnen Methoden wird im Detail diskutiert. Basierend auf diesen Ergebnissen wende ich die Methoden auf echte Datensätze elektrochemischer Experimente sowie paläoklimatischer Proxydaten an um Regimeübergänge zu detektieren. Eine herausgehobene Stellung nimmt hier die Anwendung auf den ostasiatischen und indonesisch- australischen Sommermonsun ein, jedoch gibt es noch keine Erklärung für detektierte Wechselwirkungen im tausendjährigen Rhythmus. Ein Verständnis der Phasenlage zwischen den System ist notwendig um Erkenntnisse über die langfristige Variabilität des erweiterten ostasiatisch-indonesisch-australischen Monsunregimes zu erlangen. Hier wende ich die neuen Methoden auf Speläothem-Proxydaten aus dem Holozän an, die in Nordwestaustralien und Südchina aufgenommen wurden. Dabei zeigt sich eine bipolare, hin-

und herschaukelnde Wechselbeziehung zwischen den regionalen Monsunsystemen. Diese antiphasige Beziehung kann mit Verschiebungen der innertropischen Konvergenzzone durch die Sonnen-Variabilität und damit verbundenen tropischen Niederschlägen in Verbindung gebracht werden.

Die vorgestellten Ansätze bieten geeignete Methoden um dynamische Übergänge zu untersuchen. Sie können unterschiedlichste Anwendungen in einer Vielzahl von Zeitreihenanalysen finden. Beispielsweise wird generell die Untersuchung irregulär gesampelter Datensätze ermöglicht. Letztlich, da die Zusammenwirkung der Proxydaten aus Nordwestaustralien und Südchina besser verstanden ist, stellt sich die Aufgabe solche Analysen mit Hilfe der hier vorgestellten Methoden auf Paläoklimadaten der gesamten Erde auszuweiten, um ein umfassendes Netzwerk der paarweisen Wechselbeziehungen zu erstellen, das für ein besseres Verständnis des Erdsystems hilfreich sein kann.

List of publications

- Eroglu, D., Marwan, N., Prasad, S. and Kurths, J.: Finding recurrence networks' threshold adaptively for a specific time series, Nonlin. Processes Geophys., 21, 1085-1092, doi:10.5194/npg-21-1085-2014, 2014.
- Eroglu, D., Peron T. K. DM., Marwan N., Rodrigues F. A., Costa L. da F., Sebek M., Kiss I. Z. and Kurths J.: Entropy of weighted recurrence plots, Phys. Rev. E 90(4), 042919, doi:10.1103/PhysRevE.90.042919, 2014.
- Ozken I., Eroglu D., Stemler T., Marwan N., Bagci G. B. and Kurths J., Transformation-cost time-series method for analyzing irregularly sampled data, Phys. Rev. E 91, 062911, doi:10.1103/PhysRevE.91.062911, 2015.
- Afsar O. , Eroglu D., Marwan N., and Kurths J., Scaling behaviour for recurrence based measures at the edge of chaos, EPL, 112 10005, doi:10.1209/0295-5075/112/10005, 2015.
- Eroglu D., McRobie F.H., Ozken I., Stemler T., Wyrwoll K.-H., Marwan N. and Kurths J., The Holocene East Asian–Australian summer monsoon: A see-saw relationship, (in revision), 2015.

Acknowledgements

There are many people that I have to thank for helping during my time in beautiful The Albert Einstein Science Park, Telegrafenberg. Most importantly, I would like to thank Prof. Dr. Jürgen Kurths both for his scientific guidance as well as for showing me how to become a great person. I am grateful to all his patience, encouragement, and support.

A wonderful friend, teacher, motivation and scientific problem source, Norbert Marwan. This thesis would not become true without his great help. I am deeply grateful to him.

I can't thank my mates enough, for perfect friendship and amazing help for building this work. First, although there are thousands of kilometres between us, I always have had Thomas Stemler's company, help and jokes. And I am thankful to him for giving me opportunities to visit the Overseas and perfect time during my stay whether they've taken place in the office or the pub. Second, my first officemate, Paul Schultz, it was not possible to survive in Potsdam without his support! Next, a good friend of long Skype talks and perfect times as two foreigners in Germany, Thomas Peron. The most difficult challenger of FIFA, Carsten Grabow (aka Dr. Grabowski), I thank him for memorable discussions at the island with our beers. Of course, the tallest, Julian Maluck, not just being a perfect friend but also eating all the left overs!

I also have to thank my old friends from my former university, Ibrahim Ozken (aka Ibo) and Ozgur Afsar, for joining me in Potsdam and having great time abroad. I also thank Ugur Tirnakli for his friendship and unlimited support in academical and social life. I also thank my past, present and future friend/collaborator Tiago Pereira who helped me at all steps of my scientific life. I also would like to thank my former supervisor Baris Bagci who had inspired me to come Potsdam.

I also would like to thank all group of Prof. Dr. Kurths– in particular, Jojo, Chiranjit, Niklas, Aljoscha, Bedertha, Marc, Frank, Veronika, Lyuba, Reik, Jonathan, Jasper, Jonatan, Tim, Sabine, Jobst, Jakob, Nora, Kira, Yang, Dong, Maik, Peng, Peter. I also wish to thank the Potsdam-Institut für Klimafolgenforschung (PIK) for its remarkable research environment, for its astonishing campus. I also would like to thank Gabi, Anja, Till, Heike and Nadine for being the cornerstones of solving administrative things.

I also acknowledge the Leibniz Association (WGL) under Grant No. SAW-2013-IZW-2 for financial support.

Finally I would like to thank my parents for their unconditional trust, support and love.

Special thank goes to my wife Merve for being patient of me. Without her patience, belief and love this would not be accomplished. Thank you very much my life source.

Contents

List of Figures

List of Tables

List of frequently used abbreviations

Mathematical symbols

$\lvert \cdot \rvert$	absolute value
$\lVert \cdot \rVert$	norm
$\{u_i\}_{i=1}^{N}$	set with N elements
$\boldsymbol{R}(\epsilon)$	recurrence matrix
ε	recurrence threshold
ℓ	diagonal line length
ν	vertical line length
$\Theta(\cdot)$	Heaviside function
τ	time delay for phase space reconstruction
$\boldsymbol{A}$	adjacency matrix
BC	betweenness centrality
DET	measure of recurrence quantification analysis: determinism
L	measure of recurrence quantification analysis: average diagonal length
$\boldsymbol{L}$	Laplacian matrix
m	embedding dimension for phase space reconstruction
$P(\cdot)$	histogram or frequency function
$p(\cdot)$	probability distribution
$\mathbb{R}$	set of real numbers
S_x	Shannon entropy of x
s_i	strength of ith time step for a recurrence plot
T	transitivity
$\widetilde{\mathbf{W}}$	weight matrix

Abbreviations

EAIAM	East Asian–Indonesian–Australian monsoon
EAM	East Asian monsoon
IAM	Indonesian–Australian monsoon
ITCZ	intertropical convergence zone
RN	recurrence network
RP	recurrence plot
RQA	recurrence quantification analysis
ST	supertrack functions
TACTS	transformation cost time series
wRP	weighted recurrence plot

Chapter 1

Introduction

Mankind has observed and still is observing many interesting sequences of natural events such as climate change, earthquakes, biological and chemical activities etc. In order to investigate these interesting complex structures and predict future system behaviour, many methods have been introduced and continuously improved, for instance modelling physical phenomena with mathematical objects, creating experimental setups for laboratory environments, performing computer simulations and measurement techniques themselves. In the beginning of physical studies, physicists have centered upon classical (linear and predictable) systems, but in reality all these systems usually have linear and nonlinear properties. Exploration of natural complex systems has shown that they are, in general, nonlinear in their dependence of interaction between the elements and the velocities. For example, the fundamental equations of fluid dynamics (the Navier-Stokes equations (Acheson, 1990)), oscillating circuits (Liénard equation (Liénard, 1928)) or similar mathematical models are nonlinear. The theory of nonlinear dynamics or dynamical systems consists of techniques for analyzing iterated maps or flows (differential equations). This mathematical theory is one way to analyze natural developments. The qualitative theory of nonlinear dynamics is rooted to a great extent in the seminal work of Henri Poincaré in 1890 (Poincaré, 1890). Poincaré presented a method which provides a local as well as a global analysis of nonlinear differential equations. The Poincaré recurrence theory and stability theory for fixed points and periodic orbits are included in his work(Poincaré, 1890).

Due to varying conditions, nonlinear dynamics usually exhibit transitions between dynamical regimes. These transitions are a fundamental property in prototypical models and they have been found in nature and experimental setups as well. Examples include collection of thermodynamic systems (Huang, 1987), cardiac cells (Chialvo et al., 1990; Small et al., 2003), semiconductors (Ye et al., 1993), lasers (Scott et al., 1975), RLC circuits (Linsay, 1981). The prototypical models are defined via mathematical descriptions, in other words, the equation set of the system is known and dynamical regime of the system can be found via mathematical analysis. Although a complexity measure, called the Lyapunov exponent, of a dynamical system can accurately quantify the behaviour from the equations of motion, it is not easy to evaluate it in real-world examples. For real-world applications, we are just able to reach time series of physical systems. In order to characterize the behaviour of a system and detect the transitions without equations, we need to use time series analysis techniques.

Hence, time series analysis is a very important research field in science. In the beginning, data analysis started with linear techniques as well. Obviously more realistic and advanced nonlinear methods were soon introduced and started the field of nonlinear time series analysis (Kantz and Schreiber, 1997). In general, these techniques use the fundamental features of the nonlinear dynamics such as state properties, evolution rule (deterministic or stochastic), complexity etc. for uncover underlying dynamics (Kantz and Schreiber, 1997).

Recurrence plot (RP) is one famous time series analysis technique. The main idea behind the RP is the Poincaré recurrences and the RP tool reflects the revisits of trajectories in phase space to a recurrence matrix. Thus, one can analyze the matrix and uncover the related dynamical features in the system, but RP is not applicable for all cases.

Analyzing time series from real-world applications has some crucial difficulties such as noise, cumulative trends, non-stationary situations and irregularities in the sampling. These difficulties can be found in several disciplines like neuroscience (Hensman et al., 2013), physiology (Kreindler and Lumsden, 2012), astrophysics (Thiebaut and Roques, 2005) and earth sciences (Rehfeld and Kurths, 2014). In order to apply RP on these kind of problematic time series, first one needs to apply preprocessing procedures. Obviously, the most optimal case for the preprocessing is the one that causes the lowest deformation on time series. However regularly interpolation as a preprocessing step is often applied, but this might lead to a big bias of the results (Rehfeld et al., 2011; Rehfeld and Kurths, 2014).

Applications on paleoclimate data sets will be one of the main focus in this thesis. Since paleoclimate proxies provide significant truths on the system Earth and are very important to further our understanding of the past. This knowledge from the past can potentially lead to the creation of improved climate models that can better forecast future climate change and its impacts as well as increase our understanding of the climate dynamics. My particular interest is Asia-Australian monsoon activity since nearly half of the Earth's population live in this region a fact that emphases the pressing need for a full understanding of the monsoon regime in a world of changing climates.

In this PhD thesis I will present improvements on recurrence based time series analysis and how to deal with basic measuring difficulties in real data sets. One of the challenges in time series analysis is to detect dynamical changes in the evolution of the underlying system. The main focus of this thesis will be how to detect such transitions. Several applications will be given in all chapters especially in paradigmatic models and paleoclimate time series. The corner stone definitions, well-known techniques, methodological background of nonlinear dynamics will be given in the second chapter. The concept of the recurrence plot will be demonstrated and applied on simple population model. In order to uncover the transitions, the results from RP will be compare with traditional measures, namely the Lyapunov exponent and supertrack functions. The following chapters of the second one will assign to my novel studies. In the third chapter, the main measure will be the entropy of recurrence based tools. In order to get into details, I will briefly discuss what is entropy and expectations

from the entropy. RP is a limited matrix and this limitation brings some problems for traditional entropy of RP. I will introduce a new entropy of RP, which ignores the problem from the borders of the matrix. Then, I will present a new recurrence based tool, which is called weighted recurrence plot. One of the most important feature of the new tool is that it neglects an arbitrary parameter (closeness threshold) of RP. All entropy measures will be compared for two prototypical models and a real experiment data sets. The fourth chapter is devoted to network approaches in the field of time series analysis. As a theoretical background, the graph theory is briefly expressed. Then optimal creation of a recurrence network (RN) is discussed. In the application part of the chapter, regular RN and optimal RN are compared on synthetic time series as well as a paleoclimate time series from Lisan lake. In chapter five, I will discuss how to deal with analyzing real-world data sets, since in general they have some difficulties due to measurement errors. A new preprocessing technique, called TrAnsformation-Cost Time-Series (TACTS) method, will be introduced. The ability of combination of this new preprocessing technique and RP analysis will be shown in pragmatic models and a paleoclimate data set. As a last application in paleoclimate, the new method will enable us to: (i) identify monsoon regime changes at the millennial to sub-centennial time scales in East Asian and Australian summer monsoon proxy records; (ii) demonstrate that the relationship over some ~ 9000 years has taken on a seesaw relationship – with millennial scale monsoon states, essentially opposingly phased; and (iii) show that solar variability acts as a mechanism that can account for this anti- phasing. Finally I state a conclusion of all this PhD thesis.

The main findings in this thesis have been published or submitted. Particularly Chapter 2 based on (Eroglu et al., 2014b), Chapter 4 based on (Eroglu et al., 2014a) and Chapter 5 based on (Ozken et al., 2015) and a recently submitted manuscript (Eroglu et al., 2015).

Chapter 2

Nonlinear Dynamics and Recurrence Plot

Recurrence analysis is the main technique of this PhD thesis. Recurrence based nonlinear data analysis is based on the study of phase space trajectories. In this chapter, I will briefly introduce the concept of nonlinear dynamics and methods to analyze the complex geometry of their time evolution. Then, I will discuss how trajectories get close to past positions in phase space and what one can make of these recurrences.

2.1 Nonlinear Dynamics

Nonlinear dynamics is a theory that allows us to describe systems by a set of quantitative variables. These evolve in its own phase (state) space with respect to time and the given rules or equations. It originated in the work of Newton in the 17th century, when he introduced differential calculus. Using his technique, he explored the laws of motion and gravity. Moreover, Newton solved the two body problem, interacted motion of Earth and the Sun, and opened a new epoch for subsequent generations of mathematicians and physicists. Although plenty of scientists tried to improve Newton's techniques to solve the three-body problem (the Moon included into the two body problem) it remained unsolved until Poincaré's seminal work in 1890 (Poincaré, 1890). In order to solve the problem, Poincaré changed his view to the question. Instead of looking for the exact positions, he focused on the stability of orbits of planets. This approach motivated the modern theory of nonlinear dynamics.

Nonlinear dynamics can be deterministic or stochastic (random) and it can be continuous or discrete in time. Many real-world systems or paradigmatic models can be modelled by nonlinear dynamics. The motion of a pendulum, prey-predator dynamics, flow of water in a tube and many other systems can be given as examples.

2.1.1 Trajectory and Phase Space

In the continuous case, nonlinear dynamics described by d-variables are given by a set of differential equations,

$$\dot{\boldsymbol{x}}(t) = \frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}(t)), \quad \boldsymbol{f} : \mathbb{R}^d \to \mathbb{R}^d \tag{2.1}$$

where the vector field $\boldsymbol{f}(\boldsymbol{x}(t))$ is a specified autonomous function of the vectors $\boldsymbol{x}(t)$,

$$\boldsymbol{x}(t) = (x_1(t), x_2(t), ..., x_d(t))^T. \tag{2.2}$$

Typically each dynamical system starts from a given initial condition $\boldsymbol{x}_0 = \boldsymbol{x}(t = 0)$ at $t = 0$ and evolves in its phase space. For example, assume an undamped simple pendulum,

$$\ddot{\boldsymbol{x}} + \frac{g}{\ell} \sin(\boldsymbol{x}) = 0, \tag{2.3}$$

where $\boldsymbol{x}$ is the angle of the pendulum, g is the gravity and ℓ is the length of the pendulum Fig. 2.1(a).

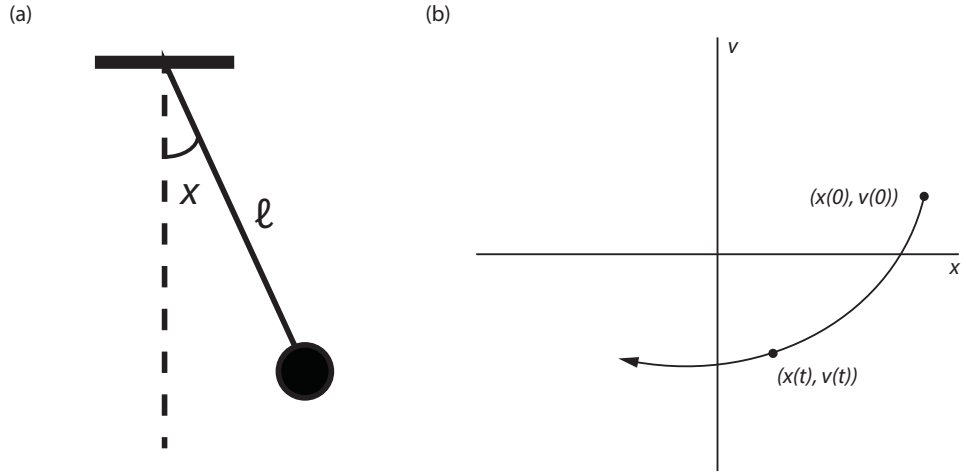

Figure 2.1: (a) Illustration of a simple pendulum, (b)trajectory of a simple pendulum for a particular initial condition in the pendulum's phase space.

The equivalent system can be rewritten in expanded and nonlinear form,

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= -\frac{g}{\ell} \sin(x). \end{aligned} \tag{2.4}$$

Where x and v represent the angular position and the angular velocity of the pendulum respectively. The solution of the pendulum for a given initial condition $x(t = 0)$ and $v(t = 0)$ is shown in Fig. 2.1(b). In Fig. 2.1(b), the curve is called trajectory (or orbit) and the space we visualize it in is called the phase (or state) space of the system.

2.1.2 Typical behaviour of Nonlinear Dynamics and the Lyapunov Exponent

Another very important feature of the system is the behaviour of the trajectories. There are three main types of behaviour: periodic, quasi-periodic and chaotic dynamics. The periodic behaviour is the evolution of the trajectories that repeats its values in regular intervals or periods ($x(t) = x(t+T)$, where T is the period of the system). The quasi-periodic behaviour has some similarities with the periodic one, but their behaviour is not exactly the same. Let me explain the quasi-periodic behaviour with an example: Imagine a torus-shaped phase space. The trajectory is moving on the torus with constant turning frequencies on two different axes but the trajectory never returns to one of the the previous positions.

Chaos is the behaviour of nonlinear dynamics that are (i) strictly dependent on the initial conditions, (ii) are topologically mixing and (iii) have dense periodic orbits. Although the dynamical system is deterministic, indistinguishable differences in initial conditions can bring largely diverging trajectories. Which means, the initial conditions determine the future of the system and it is almost impossible to predict the motion of the trajectories for long-term periods.

A complexity measure, called *the Lyapunov exponent*, detects and quantifies the behaviour of a dynamical system, in other words, the Lyapunov exponent characterizes the rate of divergence of infinitesimally close trajectories. It is easy to calculate the distance between the trajectories,

$$\delta(t=0) = \delta_0 = \|\boldsymbol{f}(\boldsymbol{x}_0) - \boldsymbol{f}(\boldsymbol{x}_0 + \epsilon)\|, \tag{2.5}$$

where δ_0 is the distance between the trajectories at the initial state ($t = 0$) and ϵ is very small mismatch parameter. For chaotic behaviour, it is well known that the distance between the trajectories increases exponentially while $t \to \infty$,

$$\|\delta_t\| \sim e^{\lambda t}\|\delta_0\|, \tag{2.6}$$

where λ is the Lyapunov exponent. For d-dimensional systems, there is a spectrum of the Lyapunov exponents and the number of the elements in the spectrum is equal to the dimensionality d of the phase space. In order to find the behaviour of dynamical system, one can consider the largest value of the spectrum, called maximum Lyapunov exponent, since it determines the dominant behaviour. If the maximum Lyapunov exponent is positive ($\lambda > 0$), the system is chaotic, otherwise ($\lambda < 0$), the behaviour of the system can be periodic or quasi-periodic.

2.1.3 Poincaré Recurrence Theorem

Recurrence is one of the fundamental properties of nonlinear dynamics. As I have mentioned before, Poincaré has contributed a lot to the theory of nonlinear dynamics. One of his contribution is known as the *Poincáre Recurrence Theorem* (Poincaré, 1890). The theorem states that almost all trajectories of nonlinear dynamics will return very close to their initial positions after a sufficiently long but finite time. In

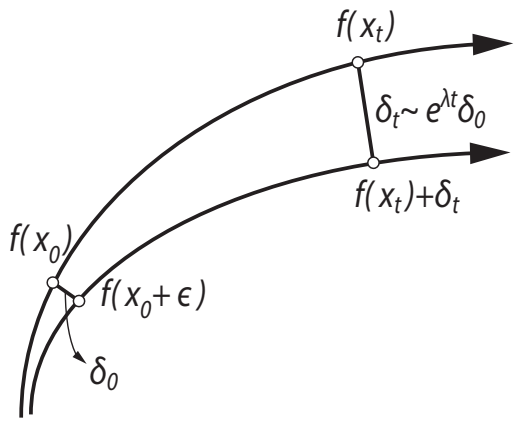

Figure 2.2: Exponential divergence of two trajectories of a chaotic system.

more detail, the theorem is the following (Katok and Hasselblatt, 1995; Barreira, 2006): Let (X, Σ, μ) be a probability space where X is the sample space, Σ is σ-algebra which is a collection of all the events and μ is the probability measure which is a function returning an event's probability. Let $T : X \to X$ be a measure-preserving transformation. Then for any set $E \in \Sigma$ and infinitely many positive n's,

$$\mu(\{x \in E : T^n x \notin E\}_{n \geq 1}) = 0. \tag{2.7}$$

In other words, there are almost no x points in the set E that will not visit the set E in finite time.

Although the theorem does not tell anything about time intervals between recurrences, it guarantees the existence of recurrence. But return times are very important dynamical features as well. Considering the same system as in Poincaré theorem, (X, Σ, μ, T) is a finite probability space and let $T : X \to X$ be a measure-preserving transformation. Then the first return time of the set E is given by

$$\tau(E) = min\{n > 0 : T^n E \cap E \neq \varnothing\}. \tag{2.8}$$

Statistical studies on the recurrences and the first recurrence times show that there are strong connections with basic characteristics of nonlinear dynamics (Kac, 1947; Hirata et al., 1999; Penné et al., 1999), such as Pesin's dimension (Afraimovich, 1997), the point-wise and local dimensions (Afraimovich et al., 2000; Afraimovich et al., 2003; Gao, 1999), the Hausdorf dimensions (Barreira and Saussol, 2001) or multi-fractal properties (Hadyn et al., 2002; Saussol and Wu, 2003).

So far I have briefly discussed two main concepts: the Lyapunov exponent as a measure to determine the behaviour of the dynamical system and the recurrences

as fundamental characteristic of nonlinear dynamics. Unsurprisingly, as a result of return time statistics as well, there is a relation between these two measures and the lower boundary of the Lyapunov exponent is given by

$$\lambda \geq \left(\lim_{r \to 0} \frac{\tau(B_r(x))}{-\log r} \right)^{-1}, \tag{2.9}$$

where $B_r(x)$ is a ball of radius r around state x and τ is the function of the first return time. (Saussol et al., 2002; Saussol et al., 2003).

2.2 Recurrence Plot

Among the different approaches to investigate dynamical properties by recurrence, the recurrence plot (RP) is a multifaceted and powerful approach to study different aspects of nonlinear dynamics. Firstly conceived to visualize time dependent behaviour of complex nonlinear dynamics, RP have been shown to be a powerful technique to uncover statistically many characteristic properties of such systems (Eckmann et al., 1987; Marwan et al., 2007).

2.2.1 Definition

In a given m-dimensional phase space, two points are considered to be recurrent if their state vectors lie in a neighbourhood characterized by a threshold ε. Formally, for a given trajectory $\mathbf{x}_i$ ($i = 1, \ldots, N, \mathbf{x} \in \mathbb{R}^m$), the recurrence matrix $\mathbf{R}$ is defined as

$$R_{i,j}(\varepsilon) = \Theta(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|),\ i, j = 1, \ldots, N, \tag{2.10}$$

where N is the trajectory length, $\Theta(\cdot)$ is the Heaviside function, and $\|\cdot\|$ is a norm of the adopted phase space (Marwan et al., 2007). The norm can be selected in different shapes: circle (Euclidian), square (maximum or supremum) or square diamond. The maximum norm is often applied, since it is computationally fast and more convenient for analytical studies. In a recurrence plot, elements $R_{i,j} \equiv 1$ (recurrence) are usually said to be black dots, whereas $R_{i,j} \equiv 0$ (no recurrence) are usually called white dots (Fig. 2.3). The original phase space can be reconstructed via time delay embedding for a time series $\{u_i\}_{i=1}^N$ (Packard et al., 1980)

$$\mathbf{x}_i = (u_i, u_{i+\tau}, ..., u_{i+\tau(m-1)}), \tag{2.11}$$

where m is the embedding dimension and τ is the embedding delay. The dimension m can be found by the false nearest neighbours method (Kennel et al., 1992) and the delay τ by the mutual information (MI) or the auto-correlation function (ACF) (Kantz and Schreiber, 1997).

The main diagonal of $\mathbf{R}$, $R_{i,i} \equiv 1$, $\forall i$, shows the line of identity (LOI). The RP is a symmetric and binary matrix. The motifs of line segments in a RP occur according

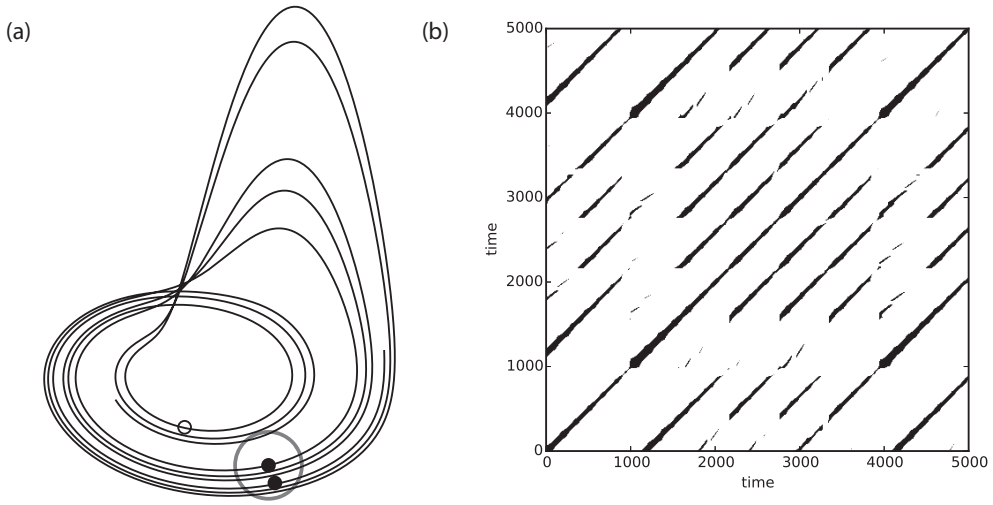

Figure 2.3: (a)Phase space trajectory of a Rössler system,$(\dot{x}, \dot{y}, \dot{z})^T = (-y - z, x + ay, b + z(x - c))^T$, integrated for 5000 time steps, with $a = 0.2, b = 0.1$ and $c = 5.7$, (b) its corresponding recurrence plot by using its three phase components and the neighbourhood threshold $\varepsilon = 2.5$ (grey circle in (a)); the Euclidean norm is used. Three points of the trajectory are highlighted with two black and one white points. The neighbourhood is determined by ε-threshold for one of the black points at time i, the other black point at time j falls into the circle. Therefore, due to symmetry, these two points are recurrent to each other and the positions $R_{i,j}$ and $R_{j,i}$ in the recurrence plot are marked with a black dot. The white point at time k is not in the neighbourhood, the positions $R_{i,k}$ and $R_{k,i}$ are coloured with white dots.

to the dynamical patterns of the underlying system. If the dynamics is given by uniformly distributed white noise (e.g. Fig. 2.4(a)), a homogeneous distribution of black points is observed. If the system is deterministic, the matrix displays diagonal line segments of black dots (Fig. 2.4(a,b)). The length of these diagonals is related to the divergence of the trajectories and is associated with the dynamics of the system. Due to this intrinsic relationship between the system's dynamics and the distribution of line segments in RPs, measures of complexity based on line segments have been introduced in order to study the dynamical properties of different systems (Trulla et al., 1996; Marwan et al., 2007).

The threshold ε is a crucial parameter. In the definition of the Poincaré recurrence theorem, the threshold is declared just as a sufficiently small distance in the system's phase space. But the selection of very small ε can lead to almost no recurrence points and of course, no reasonable RPs. On the contrary, if ε is chosen very large, almost all elements of the recurrence matrix are assigned to one and an analysis can be meaningless. Therefore, arises the need to find a proper value of ε.

In the literature, several ways of threshold selection techniques are suggested, for example ε is chosen regarding to certain percentage of the maximum phase space

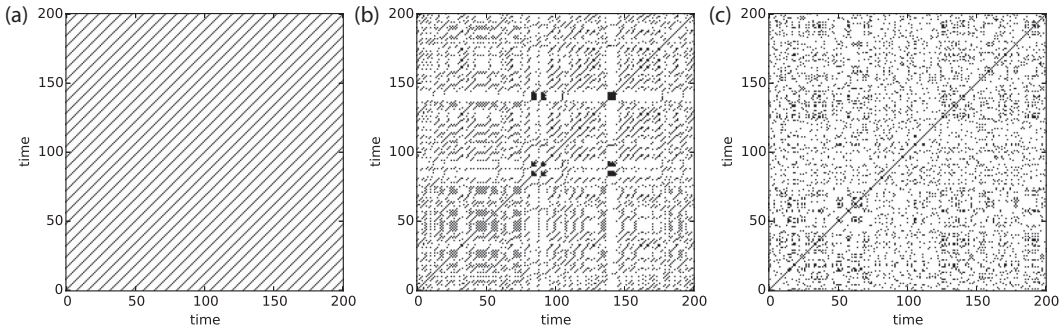

Figure 2.4: Various RPs for different nonlinear dynamics for 200 time steps: (a) periodic case for the logistic map, $x_{t+1} = rx_t(1 - x_t)$ system with $r = 3.5$, (b) chaotic case for the logistic map, $r = 4.0$, (c) uniform random distribution.

diameter (Mindlin and Gilmore, 1992; Koebbe and Mayer-Kress, 1992; Zbilut and Webber, 1992).

In real-world applications (real signals) or artificial data with a measurement noise component, the phase space is not fixed. In order to standardize the threshold distance ε for different systems, the standard deviation of the signal is used (Thiel et al., 2002).

Recently, Schinkel *et al.* have introduced a new way to standardize the threshold ε using a fixed density of recurrence points in the RPs (Schinkel et al., 2008).

In my novel studies, I have discussed this crucial threshold selection procedure in more detail and introduce a new approach to select the threshold without information loss (Eroglu et al., 2014a). This study can be found in Chap.4

2.2.2 Recurrence Quantification Analysis (RQA)

In this section, I will discuss the quantitive meaning of RPs. There are very interesting structures in RP given by the diagonal and vertical (or horizontal) lines, which all depending on dynamical properties of the systems. The meaning of these structures have been proposed with various complexity measures in (Zbilut and Webber, 1992; Webber and Zbilut, 1994; Marwan et al., 2002). These measures are based on the density of RPs or the line structures and known as recurrence quantification analysis (RQA).

As I have mentioned before, the behaviour of a dynamical system is very important feature. These RQA measures are used to detect changes in the regime of the systems, bifurcation points, similarities between different systems etc (Trulla et al., 1996; Marwan et al., 2007). Now I will present some important measures of the RQA and their meaning.

Recurrence density

A recurrence plot is a symmetric (of course square) and binary matrix. The total number of the elements in the matrix is NxN, where N is the total number of measurement points. The basic measure of RPs is the recurrence rate (RR), which is the ratio of recurrence points to the total number of elements in a RP and given as

$$RR(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}(\varepsilon). \tag{2.12}$$

Although RR is a RQA measure, it is also a well-known complexity measure called correlation dimension (Grassberger and Procaccia, 1983c; Grassberger and Procaccia, 1983b; Grassberger, 1983a; Grassberger and Procaccia, 1983a; Grassberger and Procaccia, 1984).

Phase-space binning and analyzing the density information of the bins is a standard method to measure the complexity of a dynamical system. Methods such as Shannon entropy, Hausdorff dimension, the Kolmogorov complexity etc, can be used on the density information to quantify transitions between different dynamical regimes. In order to create a recurrence plot, we divide the phase space into equi-distant but overlapping bins. The information in the bins shows us similar features as in the measures mentioned before. In Chap.3, I will focus on the density information of recurrence plot bins and propose a new complexity measure.

Diagonal lines

Measures based on the length ℓ of diagonals in recurrence matrices ($R_{i,j} = 1$) are often used to quantify the complexity of a given RP. The frequency distribution of diagonal line lengths $P(\ell, \varepsilon)$ is directly linked with the dynamics, hence related with the Lyapunov exponent, since$P(\ell, \varepsilon)$ quantifies the divergence behaviour of the dynamical system(Trulla et al., 1996; Thiel et al., 2004a; Marwan et al., 2007).

$$P(\ell, \varepsilon) = \sum_{i,j}^{N-\ell} (1 - R_{i-1,j-1}(\varepsilon))(1 - R_{i+\ell,j+\ell}(\varepsilon)) \prod_{k=0}^{\ell-1} R_{i+k,j+k(\varepsilon)}. \tag{2.13}$$

For simplicity, I discard the symbol ε in the histogram of diagonal line lengths (i.e. $P(\ell, \varepsilon) = P(\ell)$) and relatedly I suppress the dependence on ϵ in the following, as all measures depend on it anyway.

One of the most important measures of RQA is the determinism (DET) which is quantifying the fraction of recurrence points $R_{i,j} \equiv 1$, that form diagonal lines (Marwan et al., 2007),

$$DET = \frac{\sum_{\ell=\ell_{\min}}^{N} \ell P(\ell)}{\sum_{\ell=1}^{N} \ell P(\ell)}, \tag{2.14}$$

where $\ell_{\min}$ is the consideration of diagonal lengths. For $\ell_{\min} = 1$, the determinism is one and for larger selection of $\ell_{\min}$, the determinism decreases. The common selection of $\ell_{\min} = 2$. The determinism DET refers to the predictability of the future of a dynamical system: high determinism for highly predictable systems and low determinism for poorly predictable systems.

Another diagonal length based measure is the average diagonal line length,

$$L = \frac{\sum_{\ell=\ell_{\min}}^{N} \ell P(\ell)}{\sum_{\ell=\ell_{\min}}^{N} P(\ell)}. \tag{2.15}$$

The average diagonal line length L is the mean time that two segments of the trajectory accompany each other in the same neighbourhood without diverging from each other.

The length of the longest diagonal line is a RQA measure as well,

$$L_{\max} = \ell_{\max} = \max(\{\ell_i\}_{i=1}^{N_\ell}), \tag{2.16}$$

where N^ℓ is the total number of diagonal lines and its inverse is the divergence measure,

$$DIV = \frac{1}{L_{\max}}. \tag{2.17}$$

The divergence DIV means that the divergence time of two trajectory segments from each other with taking into account the longest accompanied time.

Another RQA measure is the Shannon entropy,

$$S_{\mathrm{RP}} = -\sum_{\ell=\ell_{\min}}^{\ell_{\max}} p(\ell) \log p(\ell), \tag{2.18}$$

where $\ell_{\max}$ is the length of the longest diagonal line, $p(\ell) = P(\ell)/N_\ell$ is the probability of occurrence of a line of length ℓ, and N_ℓ is the total number of the line segments in the RP. S_{RP} is a heuristic complexity measure and, for this reason, does not correspond to the physical entropy of the system. Moreover, S_{RP} can show anti-correlated behaviour with the Lyapunov exponent, I will discuss these causes later in detail. This unexpected behaviour can cause misinterpretation of a system. Because of fact this, S_{RP} shouldn't be used as the Shannon entropy directly. Therefore, in order to elucidate this contradiction between the Lyapunov exponents and recurrence based entropy measurements, I have developed another recurrence based entropy definition that I will introduce in Chap.3 alongside a discussion of this paradox.

Among all RQA measures based on diagonal structures , the most common ones are introduced. In this thesis, I will use some of these RQA measures to identify the underlying dynamics of complex systems.

Vertical lines

Another interesting structure in RPs are vertical or horizontal lines. In this thesis, I will not extensively discuss vertical line-base measures, nevertheless it is worth giving the classical definitions and the meanings of them, since they give further insight on intermittency and laminar states of a system (Marwan et al., 2002; Marwan et al., 2007). Moreover, in the vertical lines' case white points ($R_{i,j} = 0$) have an important meaning since they possess information on recurrence times.

The frequency distribution of vertical line lengths $P(\nu)$ is given by

$$P(\nu) = \sum_{i,j}^{N-\nu} (1 - R_{i,j})(1 - R_{i,j+\nu}) \prod_{k=0}^{\nu-1} R_{i,j+k}, \tag{2.19}$$

where ν is the length of vertical lines. The RQA measures are based on the histogram of diagonal lines $P(\nu)$.

As previously mentioned, the vertical structures are related to the laminar states and therefore the first measure called laminarity, is given by

$$LAM = \frac{\sum_{\nu=\nu_{\min}}^{N} \nu P(\nu)}{\sum_{\nu=1}^{N} \nu P(\nu)}. \tag{2.20}$$

The definition of the laminarity is analogous to the determinism. LAM quantifies the fraction of recurrence points $R_{i,j} \equiv 1$, that form vertical lines and, similarly as for determinism, a common adopted choice for the minimum length of vertical lines is $\nu_{\min} = 2$(Marwan et al., 2007).

It is also possible to quantify how long the system stays in a state through the trapping time, which is defined as,

$$TT = \frac{\sum_{\nu=\nu_{\min}}^{N} \nu P(\nu)}{\sum_{\nu=\nu_{\min}}^{N} P(\nu)} \tag{2.21}$$

it is similar to the average diagonal line length. TT as a measure estimates the average trapping time that is how long the system will wait in a state.

The last measure given here is the maximal vertical line length

$$\nu_{\max} = \max(\{\nu_i\}_{i=1}^{N_\nu}), \tag{2.22}$$

similar to L_{max}, where N^ν is the number of total vertical lines.

The different RQA measures are able to specify various properties of nonlinear dynamics. For example, diagonal-based measures are very suitable to detect transitions between periodic-chaotic dynamics but vertical-line-based ones are able to distinguish chaos-chaos regime transitions (Marwan et al., 2007). Thus, using different kind of measures of RQA allows us to study a variant of dynamical properties of a system. Besides, the RP framework is a very powerful nonlinear approach to detect dynamical

properties that also provides a practical (easy) way to analyze different substance of systems.

Application: Detection of transitions via various RP measures, based on diagonal and vertical lines

As mentioned before, RP is a powerful tool to detect transitions and the logistic map is a suitable one dimensional model to investigate phase transitions between different dynamical regimes since its Lyapunov exponent and supertrack function (Oblow, 1988) are well known. The logistic map is a discrete map given by

$$x_{i+1} = rx_i(1 - x_i) \tag{2.23}$$

where x is a real number between 0 and 1, and r is the control parameter of the system. The logistic map possesses rich dynamical properties between $r \in [3.5, 4]$, e.g., periodic and chaotic dynamics, bifurcations or inner-outer crises. For each value of r, I create a time series of length $N = 5000$. In order to discard transients, I use only the last 3000 values of this time series. It was shown that RQA measures can distinguish different dynamical regimes of the system, like chaos-periodic and chaos-chaos transitions (Trulla et al., 1996; Marwan et al., 2002).

Now I calculate the Lyapunov exponent and various diagonal-based RP measures. In this special case, I do not apply a reconstruction (embedding) preprocess since the logistic map is a one-dimensional map.

Diagonal structures of RP are related with the divergence of two different trajectories, as described above. Therefore, it is expected that diagonal based measures are correlated with the Lyapunov exponent. The Lyapunov exponent is a measure to distinguish the regime of the dynamics. Thus, diagonal-based RP measures are able to detect transitions between a chaotic regime and a periodic regime.

Here I perform three diagonal line based measures, namely determinism (DET), average diagonal line length (L) and divergence (DIV), in order to investigate the transitions between different regimes. In Fig. 2.5, the measures clearly identify the transitions where the Lyapunov exponent hits or crosses the zero line (given by red straight line). When the Lyapunov exponent increases from negative part and hits the red line ($\lambda = 0$) then decreases again, there is a bifurcation point. If the curve crosses the red line from positive to negative (or other way round), there is a transition from periodic to chaotic regime (or chaotic to periodic one). The RP based analysis allows us to identify both cases (Fig. 2.5).

Vertical structures of RP possess information about intermittency and laminar states of the system.While varying the control parameter of the logistic map, due to bifurcations some chaotic bands occur. These bands converge for some critical control parameters. In order to detect these transitions, the Lyapunov exponent is not convenient anymore but supertrack functions (Oblow, 1988) of the system are able to catch these transitions.

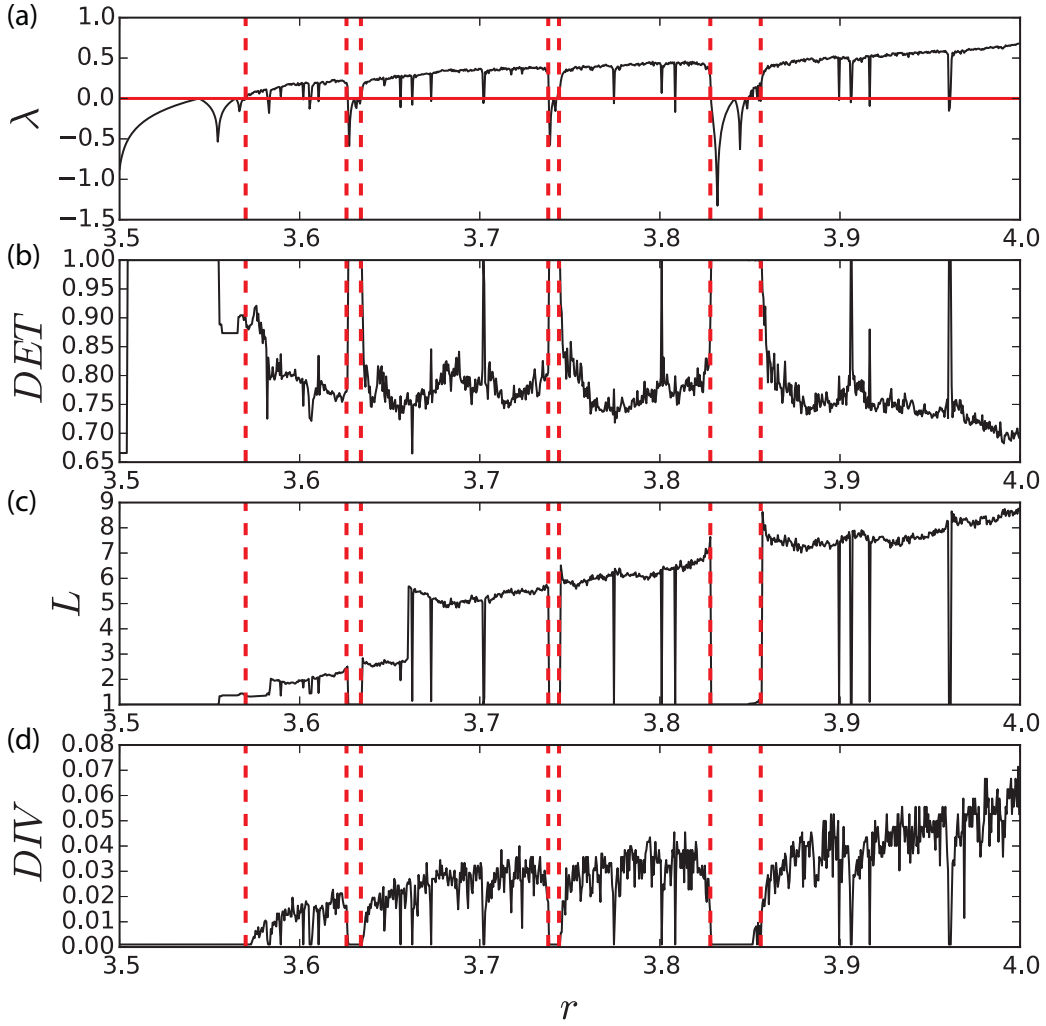

Figure 2.5: Comparison between (a) the Lyapunov exponent λ and the diagonal-line-based recurrence plot measures (b)determinism DET, (c) average diagonal line length L and (d) divergence DIV for the logistic map. The intersection points of the Lyapunov exponent and the red line ($\lambda = 0$) are given by dashed red lines, RP based measures show very sharp abrupt changes at the critical points.

In order to investigate chaos to chaos transitions, I used three vertical-line-based RP measures: laminarity (LAM), trapping time (TT) and maximal vertical line length ($\nu_{\max}$). Low-ordered supertrack function curves possess intersections. These intersections display chaotic band merging points. In real world applications, usually, there is no way to acquire a mathematical formulation of the system. Due to this reason, there is no possibility to compute supertrack functions. Thus, to have a way to distinguish chaos to chaos transition points directly from the time series of a system is very important convenience and property.

In Fig. 2.6, I plotted low-ordered supertrack functions (ST) and pointed out the band merging points with dashed red lines. As it is seen below the ST subfigure, vertical-line-based RP measures show unordinary peaks. These peaks refer to chaos-chaos transitions. This is very important property of RP and I introduced the result of vertical based measures just for illustration. In the following I will focus on transitions which is between different regimes.

After this brief introdcution to the theory of RQA , I will go through my novel studies which were done during my PhD. It is clear that the recurrences will be the main focus of the thesis, and I will apply this fundamental property of nonlinear dynamics.

In this thesis, I will concentrate on the detection of the underlaying properties of complex systems such as synthetical models (different level of complexity), laboratory observations and paleoclimate time series.

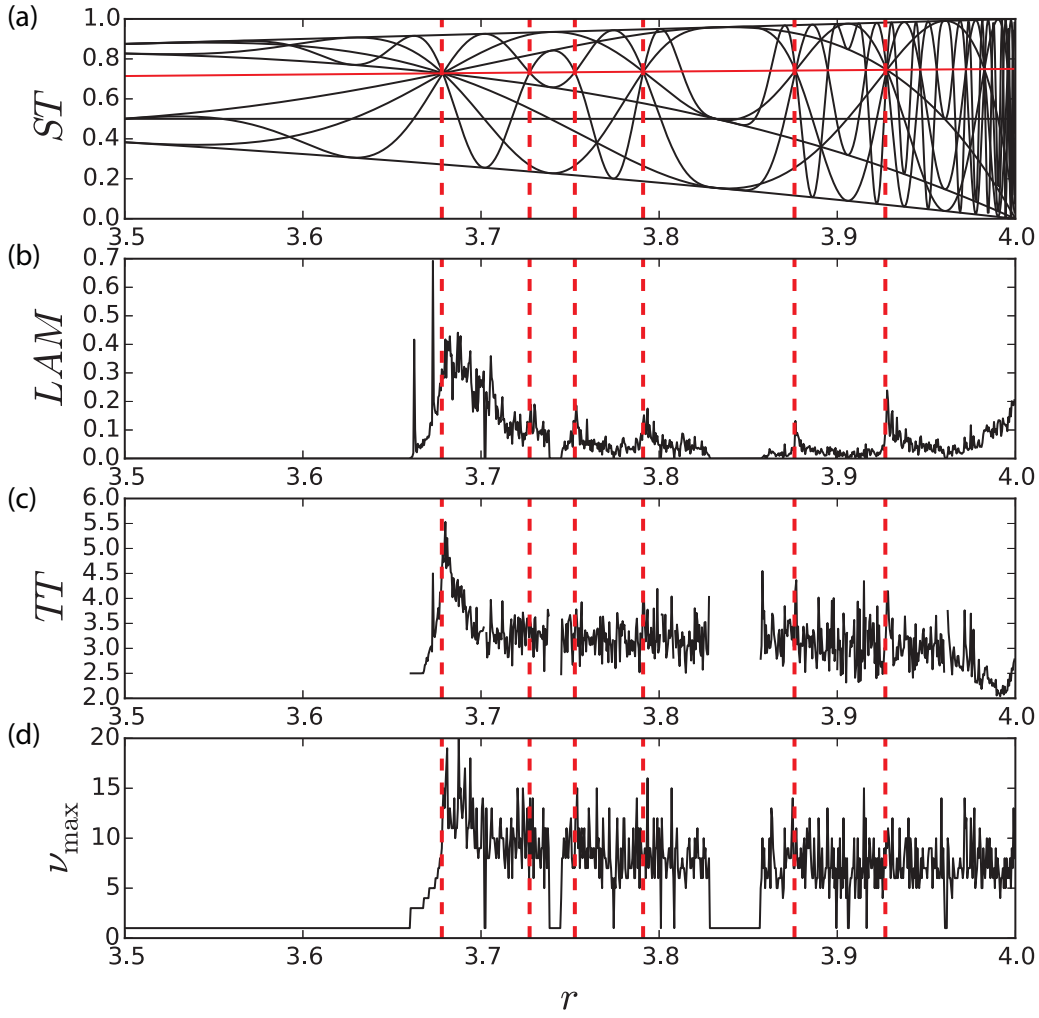

Figure 2.6: Comparison between (a) the low ordered supertrack functions ST and the vertical lines based recurrence plot measures (b) laminarity LAM, (c) trapping time T and (d) maximum vertical line length $\nu_{\max}$ for the logistic map. At the chaotic band merging points of the the low ordered supertrack functions (dashed red lines), RP based measures give clear transitions.

Chapter 3

Entropy of Regular/Weighted Recurrence Plots

The Shannon entropy of a time series is a standard measure to assess the complexity of a dynamical process and can be used to quantify transitions between different dynamical regimes. As I have mentioned in the previous chapters, recurrence quantification analysis (RQA) is an alternative way of quantifying complexity based on recurrence states in the phase space. Although various definitions for recurrence based entropies have been suggested so far, some of them show reveal inconsistent results. Here I suggest a new perspective to estimate the complexity by the Shannon entropy of the heterogeneity density of a given point in the phase space (Eroglu et al., 2014b). The associated Shannon entropy is positively correlated with the largest Lyapunov exponent.

Moreover, the threshold value ε used to define the distance of recurrence points, is one of the biggest handicap of the recurrence plot approach. In this chapter I suggest a different approach to create a distance based matrix by using a weighted variant of the recurrence plot, called weighted recurrence plot (wRP) which was published in Physical Review E, (Eroglu et al., 2014b). The wRP allows us to analyze the heterogeneity of the density of a given point in the phase space. I demonstrate the potential of wRP on a prototypical example as well as on an experimental data of a chemical experiment.

3.1 What is Entropy?

Entropy is the mean value of the information involved in each signal. In 1948, an article entitled *"A Mathematical Theory of Communication"* published by mathematician Claude E. Shannon. The idea was to use the formalism of mathematics to describe communication as a phenomenon. The entropy is introduced to measure the density of information of the messages in the communication.

In order to quantifying the complexity of a dynamical system, different entropy measures were introduced. In general, these entropy definitions are inspired by Shannon's entropy in the information theory. For instance, Sinai defined a generalized concept of entropy for measure-preserving dynamical systems, today known as Kolmogorov-Sinai-entropy (Sinai, 1959). Another entropy measure was proposed by Pincus in 1991, known as approximate entropy as a measure of system complexity

(Pincus, 1991). The last example of the entropy definitions is the K_2 entropy which is the estimation of the Kolmogorov entropy from a signal (Grassberger and Procaccia, 1983b).

Although the Shannon entropy is well-known measure, I will briefly give some details such as 'what is a fair system', 'when entropy is high or low' etc. The most convenient way to explain entropy measures, is using stochastic nonlinear dynamics, since the entropy is statistically defined. In my studies, I will use the Shannon entropy's mathematical formulation,

$$S(X) = -\sum_i p(x_i) \log p(x_i), \tag{3.1}$$

where X is discrete random variable with possible values $\{x_1, x_2, ..., x_n\}$ (Shannon, 1948).

Example–flipping a coin– Consider flipping a coin which could be fair or unfair and the possible outcomes are heads or tails. If we call the outcome probability of heads p and tails q, then the total probability equals one, $p+q = 1$. For a fair coin, $p = q = \frac{1}{2}$, otherwise $p \neq q$. For the fair case, it is clear that predicting the outcome is not easy. Since in the fair case, due to the probability distribution, outcome information is the lowest one (50% heads and 50% tails). Therefore, we have more information for the outcome of the unfair game. Related to this, when the information is high, the entropy is low Fig.3.2.

Obviously, the entropy is related with the predictability information of the outcome of a certain system. And the entropy is the highest for the fairest game, since it is the hardest case to guess the outcome. Another way to increase the entropy is increasing the complexity of the system. For example, if the outcome possibilities are increasing, the entropy is increasing as well.

Example–throwing dice– One of the simplest example to show how the outcome possibility affects the entropy is that stochastic dice thrown model. I can classify the dice regarding to number of faces. In this example, I will consider fair and n-sided dice models. The possible outcomes are $X = \{1, 2, ..., n\}$ and the probabilities are $p(1) = p(2) = ... = p(n) = p = \frac{1}{n}$. For a fair system, the entropy definition can be rewritten as,

$$S(X) = -\sum_i^n p(x_i) \log p(x_i) = -n\frac{1}{n} \log \frac{1}{n} = \log n. \tag{3.2}$$

It is easy to see from Eq. (3.2) that increasing the number of sides of a die n (number of possible outcomes), increases the entropy logarithmically (see Fig.3.2).

Summarizing these two toy examples, the complexity and the fairness of a system is directly related with the entropy of the system. This sensitivity of entropy makes itself a good measure to distinguish different systems from their different behaviours. As I have mentioned in the previous chapter, RPs have diagonal or vertical line distributions. Therefore, computing the entropy of a RP is a straight forward and

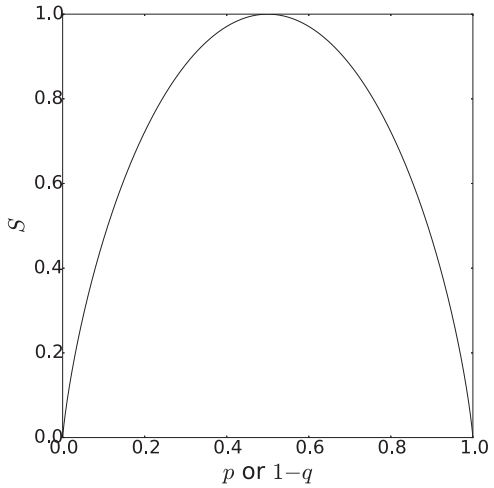

Figure 3.1: The Shannon entropy of a coin toss, the outcome possibility of heads $p = 1 - q$ and tails $q = 1 - p$. The axis of possibility can be define as the fairness of a coin. $p = q = 0.5$ for the fair game, $p \neq q$ for the unfair ones. Due to less outcome predictability, the entropy is the highest for $p = q = 0.5$. Contrary, for the unfair game, entropy is lower than fair game, since the predictability of outcome is higher, in other words we have more information to guess outcome.

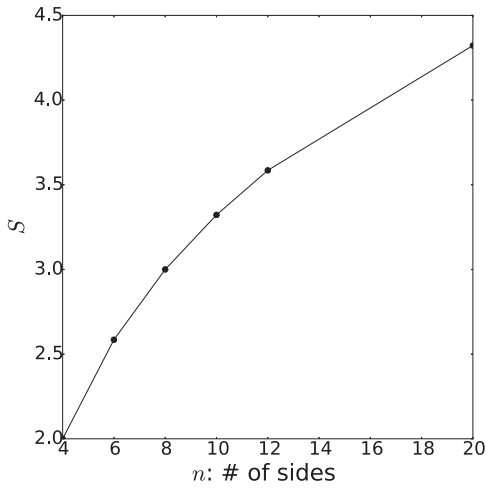

Figure 3.2: The Shannon entropy of fair n-sided dice models, for $n = \{4, 6, 8, 10, 12, 20\}$. If the number of possible outcomes n increasing for a die, entropy is increasing logarithmically too. The axis of number of sides of dice n can be represented as the complexity axis. And complexity of a system is positively correlated with the entropy.

compelling idea to analyse a dynamical system. In the next chapter, I will discuss about the pitfalls of the RP entropy and how we can overcome these.

3.2 Entropy of Recurrence Plots

Recurrence plot is a powerful nonlinear tool to analyze complex systems. There are various interesting structures in RPs, and different complexity measures are derived from them. The frequency distributions of this structures are easily convertible to probability density functions (PDF). Therefore, entropy calculations are straightforward given these PDFs. As shown in Chap.2, the entropy of RP is derived by diagonal lines of RPs,

$$S_{\mathrm{RP}} = -\sum_{\ell=\ell_{\min}}^{\ell_{\max}} p(\ell)\log p(\ell), \tag{3.3}$$

where $\ell_{\max}$ is the length of the longest diagonal line, $p(\ell) = P(\ell)/N_\ell$ is the probability of occurrence of a line of length ℓ, $P(\ell)$ is the histogram of diagonal lines, and N_ℓ is the total number of the line segments in the RP. In general, $\ell_{\min} = 2$ is an appropriate value for selecting the minimum number of diagonal length since for the larger values of $\ell_{\min}$ histogram is very sparse to analyze properly.

In RPs, expected entropy behaviour is correlated with the Lyapunov exponent, since entropy is defined by the diagonal lines and these line segments are directly linked with the underlying dynamics of the systems. However, the entropy of the diagonal line segments reveals, for some cases, a counterintuitive anticorrelation with the Lyapunov exponent, yielding high values within periodic windows and low values within chaotic regimes (Trulla et al., 1996). To solve this apparent contradiction between the notion of disorganization of a system and the value of the line length entropy (Trulla et al., 1996), another definition for a recurrence based entropy has been suggested, employing the non-recurrent (white) diagonal lines in the RP (Letellier, 2006). However, following study of us about non-recurrent diagonal lengths based entropy showed that this definition does not solve the problem (Eroglu et al., 2014a). Meanwhile, it is important to emphasize that these definitions do not correspond to the entropy of physical systems in classical statistical physics sense (Marwan, 2008).

Especially in the case of periodic nonlinear dynamics, RP presents unexpected entropy results. This unexpected behavior is the result of the following effect: When we consider, e.g., periodic dynamics, the RP should, in principle, consist of infinitely long diagonal lines. However, due to the finite time series length, these diagonal lines are cut at the borders of the RP, finally resulting in their different lengths, depending on their distance to the LOI, and hence increasing artificially the entropy measure. This biasing effect can also happen for non-periodic dynamics as long as a significant number of diagonals cross the border of the RP. For chaotic dynamics, a majority of diagonal lines are not cut by the RP border, therefore, the bias in the entropy is

negligible. Furthermore, the entropy measure also depends on the number of diagonal lines in the RP (i.e., indirectly on the recurrence threshold and the time series length), regardless of their length distribution.

In order to illustrate the frequency distribution of diagonal lengths of recurrence plot, I use the logistic map in the periodic regime. The logistic map is a simple population model and its well-known mathematical expression is,

$$x_{i+1} = rx_i(1 - x_i) \tag{3.4}$$

where x_i is a real number between zero and one that presents the ratio of existing population to the maximum one and r is a constant, in other words the control parameter of the system, $r \in [0, 4]$. For different control parameters, the logistic map shows different dynamical regimes. For this example, I will consider $r = 3.56$ which is a periodic case (period-8) for the map. Recurrence plot of the system is given in Fig.3.3(a). It is worth to note that the vertical distance between the diagonal lines is, as expected, 8 time units since the system is in period-8. In order to compute the entropy of the RP, the frequency distribution $P(\ell)$ can be found with Eq. (2.13), and the histogram is drawn in Fig.3.3(b). First, the histogram of the system looks very fair with equal possibilities but for a RP of a periodic system we expect that the diagonals are infinitely long. Therefore, the recurrence plot boundaries completely manipulate the histogram. The expected number of outcome possibility is just one which means we should have just one bar in the histogram, but instead we have a uniform distribution 3.3(b). As I have shown in the coin flip example, the uniform probability distribution function of all possible outcomes gives the possible maximum entropy for a system.

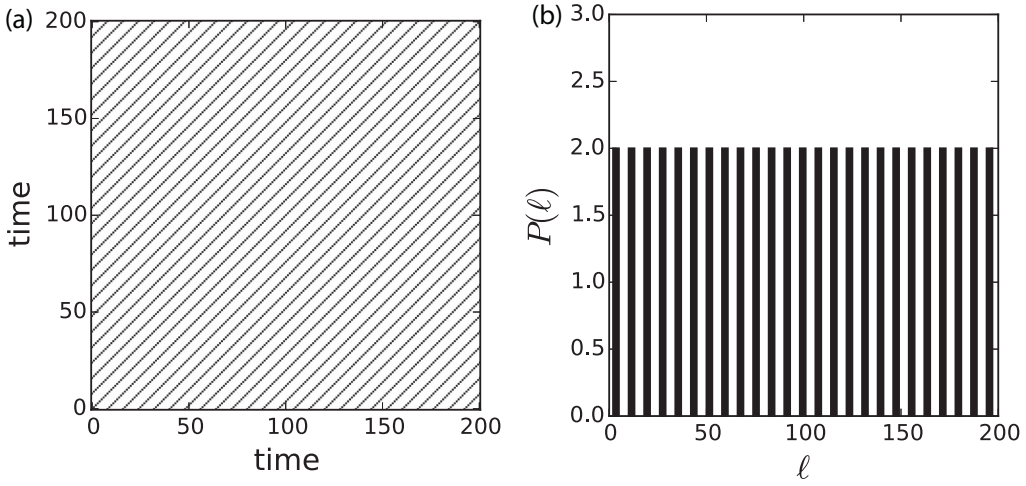

Figure 3.3: (a) Recurrence plot of logistic map for $r = 3.55$ and (b) its frequency distribution of diagonal line lengths ($P(\ell)$)

In order to get rid of the boundary effect of RPs for the diagonal line segments, one can change the shape of RP to the one shown in Fig. 3.4(a). In this case, the histogram of the diagonal line lengths is as expected (see Fig. 3.4(b)), and the entropy is zero. Although the diamond shaped RP is fair for the frequency distribution of diagonal lines, there are very large information loss and it is very unfair for the vertical line segments based measures.

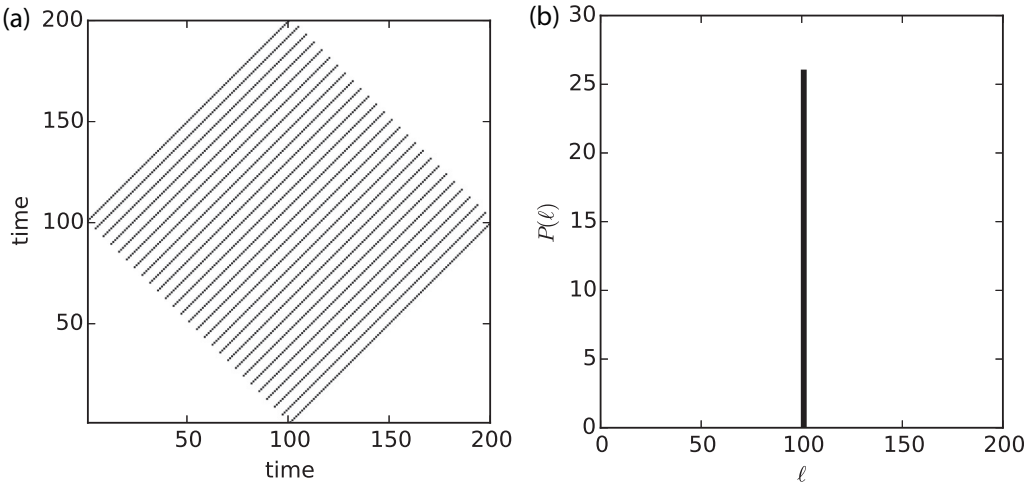

Figure 3.4: (a) Diamond shaped recurrence plot of logistic map for $r = 3.55$ and (b) its frequency distribution of diagonal line lengths ($P(\ell)$

Seeking for an alternative descriptor for the dynamical complexity to be derived from RPs, another study has suggested to calculate the Shannon entropy of the length distribution of diagonal segments of the non-recurrence points (white dots) instead of the diagonal lines formed from the recurrence points (black dots) (Letellier, 2006). In other words, let $P_{\text{white}}(\ell)$ be the number of connected diagonal non-recurrence segments. In order to compute $P_{\text{white}}(\ell)$ histogram, I define a new recurrence matrix for white points $R^{\text{white}}_{i,j}$ as,

$$R^{\text{white}}_{i,j} = 1 - R_{i,j} \tag{3.5}$$

Using the $R^{\text{white}}_{i,j}$ matrix, I can compute the histogram $P_{\text{white}}(\ell)$ of diagonal lines of length ℓ as,

$$P_{\text{white}}(\ell) = \sum_{i,j=1}^{N-1} (1 - R^{\text{white}}_{i-1,j-1})(1 - R^{\text{white}}_{i+\ell,j+\ell}) \prod_{k=0}^{\ell-1} R^{\text{white}}_{i+k,j+k}. \tag{3.6}$$

The entropy measure is then defined as (Letellier, 2006)

$$S_{\mathrm{RP}}^{\mathrm{white}} = -\sum_{\ell=\ell_{\min}}^{\ell_{\max}} p_{\mathrm{white}}(\ell) \log p_{\mathrm{white}}(\ell), \tag{3.7}$$

where $p_{\mathrm{white}}(\ell) = P_{\mathrm{white}}(\ell)/N_{\mathrm{white}}$ and N_{white} is the total number of the non-recurrence line segments in the RP. In contrast to the results presented in (Letellier, 2006), here I show that this entropy definition does not solve the problem of the anti-correlation between $S_{\mathrm{RP}}^{\mathrm{white}}$ and λ.

In addition to the entropy definitions based on the length of line segments ℓ, it is worth mentioning that we can also define entropic measures through the distribution of time returns in RPs. An example is the recurrence probability density entropy (Little et al., 2007), which measures the uncertainty associated to the distribution $P(T)$, where T is the recurrence time. Moreover, the Kolmogorov-Sinai entropy has also been generalized in the context of RPs in order to measure the complexity in the distribution of time returns (Baptista et al., 2010).

In order to provide a more intuitive RP based entropy measure, that overcomes the anti-correlation problem, I propose an entropy based on the heterogeneity of the density in the phase space that is included in the RP-matrix. The strength s_i of a point is defined in the time series by

$$s_i = \sum_{j=1}^{N} R_{i,j} \tag{3.8}$$

and the strength vector $\boldsymbol{s} = \{s_1, s_2, ..., s_N\}$ quantifying the heterogeneity of the density of each point in the phase space Fig. 3.5. Therefore, one is able to characterize the amount to statistical disorder in the system through the distribution of $P(s)$. Thus its associated Shannon entropy can be calculated by

$$S_{\mathrm{RP}}^{\mathrm{density}} = -\sum_{\{s\}} p(s) \log p(s), \tag{3.9}$$

where $p(s) = P(s)/\sum_i^N s_i$ is the probability density function of the strength vector $\boldsymbol{s}$.

The RPs have the following basic properties: (i) if the dynamics of a system is periodic, the pattern of RP is observed as continuous parallel diagonal lines along the RP (Fig. 2.4(a)). Since the RPs are limited according to the time interval of time series, different lengths of diagonals can be observed. If the values of the histogram of diagonal lines in RP are similar, the entropy is low, (see the definition of entropy Eq. (3.3)). In this way, due to the finite time of RP, different values can be observed in the histogram $P(\ell)$, which can result in deceptive results in the entropy computation, especially for periodic regimes. The next property is related to (ii) chaotic dynamics. In this case, the missing diagonal segments' length does not affect entropy measures as much as in periodic regimes. The entropy predicts higher value for chaotic regime because the values of the histograms are expected to be different from each other

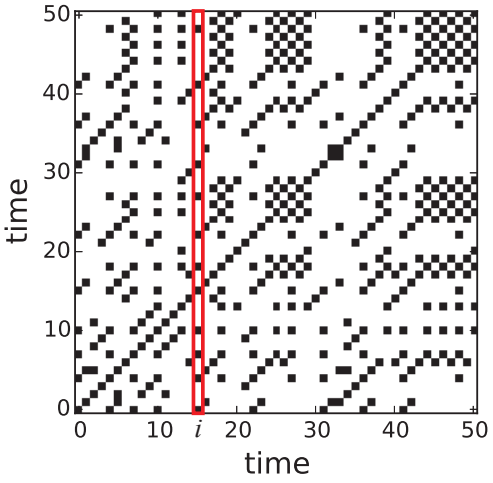

Figure 3.5: Recurrence plot of the logistic map for $r = 3.96$ in the chaotic regime. The red rectangle is the area for the neighbourhood of point i in the time series. The total number of black points in the area is the strength of the point i, $s_i = \sum_j^N R_{i,j}$.

(see Fig. 2.4(b)). Although the diagonal segments based on the histogram $P(\ell)$ are influenced from the limited time interval, the density based histogram $P(s)$ is not affected from that, because all columns (rows) of the RP are limited according to the same number of time interval.

For periodic or stochastic dynamics, the strength s_i will be very similar for all time points i, resulting in a confined distribution of $p(s)$ and finally in a very low entropy value $S_{\mathrm{RP}}^{\mathrm{density}}$. For chaotic dynamics, the strength s_i will vary strongly for different time points i, therefore the distribution $p(s)$ will be broad and $S_{\mathrm{RP}}^{\mathrm{density}}$ will increase. Thereby, it is expected that the results of the density based entropy $S_{\mathrm{RP}}^{\mathrm{density}}$ are more correlated with the Lyapunov exponent than the diagonal based entropies S_{RP}, $S_{\mathrm{RP}}^{\mathrm{white}}$.

3.2.1 Comparison of the Entropies

In this section I compare the RP based entropies using the logistic map. The logistic map is analysed within the interesting range of the control parameter $r \in [3.5, 4.0]$ with a step size of $\Delta a = 0.0005$. In this range, the logistic map shows rich dynamics, e.g., periodic and chaotic states, bifurcations, as well as interior-crisis. For each value of r, I compute a time series of length $N = 5000$. In order to discard transients, I use only the last 3000 values of this time series. It was shown that RQA measures can distinguish different dynamical behavior of the systems, like chaos-period and chaos-chaos transitions (Trulla et al., 1996; Marwan et al., 2002). As mentioned before, entropy quantifies the disorder of the system. Thus, one would expect that

with increasing chaotic nature of the system (i.e., increasing the Lyapunov exponent λ), the entropy values should increase. However, within periodic windows, the entropy should significantly decrease.

Here, I calculate the RP without embedding. For the logistic map, larger values of the embedding dimension, i.e. dimension $m = 2$ or 3 and a delay of $\tau = 1$, do not change the result significantly, because the logistic map is one-dimensional. However, the selection of the recurrence threshold ε is crucial. Hence, in order to choose an optimal threshold, I use the recurrence rate method (Marwan et al., 2007), i.e., selecting ε in such a way that the recurrence rate RR is constant at $RR = 5\%$. Sufficient small values of ε decrease the recurrence point density in RPs and lead to better distinction of small variations, whereas larger values cause denser RPs but the sensitivity to detect small variations decreases.

I compare the entropy measures S_{RP}, $S_{\mathrm{RP}}^{\mathrm{white}}$ and $S_{\mathrm{RP}}^{\mathrm{density}}$. Generally, the S_{RP} measure detects the transitions from chaotic to periodic and chaotic to periodic regimes by abrupt jumps. However, its correlation with the Lyapunov exponent λ is opposite to what one would expect. Moreover, the values within periodic windows are not consistent, e.g., for $a \in [3.50, 3.54]$ and $a \in [3.82, 3.85]$, the S_{RP} values are larger than during chaotic regimes, but for the periodic windows at $a \in [3.2...]$ or $a \in [3.75...]$, S_{RP} falls to zeros (see Fig. 3.6(b)).

Similarly, the general trend of $S_{\mathrm{RP}}^{\mathrm{white}}$ is also anti-correlated to λ (Fig. 3.6(c)). In the larger periodic windows $a \in [3.5, 3.568]$ and $a \in [3.82, 3.85]$, $S_{\mathrm{RP}}^{\mathrm{white}}$ has the highest values. However, the smaller periodic windows tend to have small $S_{\mathrm{RP}}^{\mathrm{white}}$.

On the contrary, the relationship between the Lyapunov exponent and the density based entropy $S_{\mathrm{RP}}^{\mathrm{density}}$ is generally positively correlated for the whole range of the bifurcation parameter r. For periodic states, $S_{\mathrm{RP}}^{\mathrm{density}}$ has lower values and during chaotic regions, it has larger values (Fig. 3.6(d)). At the critical values of $a = 3.544, 3.564, 3.630, 3.741$, and 3.841, $S_{\mathrm{RP}}^{\mathrm{density}}$ reveals sharp jumps.

In order to quantify the relationship between the entropies (S_{RP}, $S_{\mathrm{RP}}^{\mathrm{white}}$ and $S_{\mathrm{RP}}^{\mathrm{density}}$) and the maximal Lyapunov exponent λ, the Pearson correlation coefficient is used,

$$\rho_{xy} = \frac{cov(x, y)}{\sigma_x \sigma_y} \tag{3.10}$$

where $cov(x, y)$ is the covariance of two time series x and y, σ_x and σ_y are their standard deviations, respectively. As we can see from the Table. 3.1, S_{RP} and $S_{\mathrm{RP}}^{\mathrm{white}}$ are anti-correlated with Lyapunov exponent. On the other hand, $S_{\mathrm{RP}}^{\mathrm{density}}$ is positively correlated with the maximal Lyapunov exponent λ.

ρ_{xy}	λ
S_{RP}	-0.52
$S_{\mathrm{RP}}^{\mathrm{white}}$	-0.67
$S_{\mathrm{RP}}^{\mathrm{density}}$	0.81

Table 3.1: Correlation between entropies and Lyapunov exponent λ for the logistic map

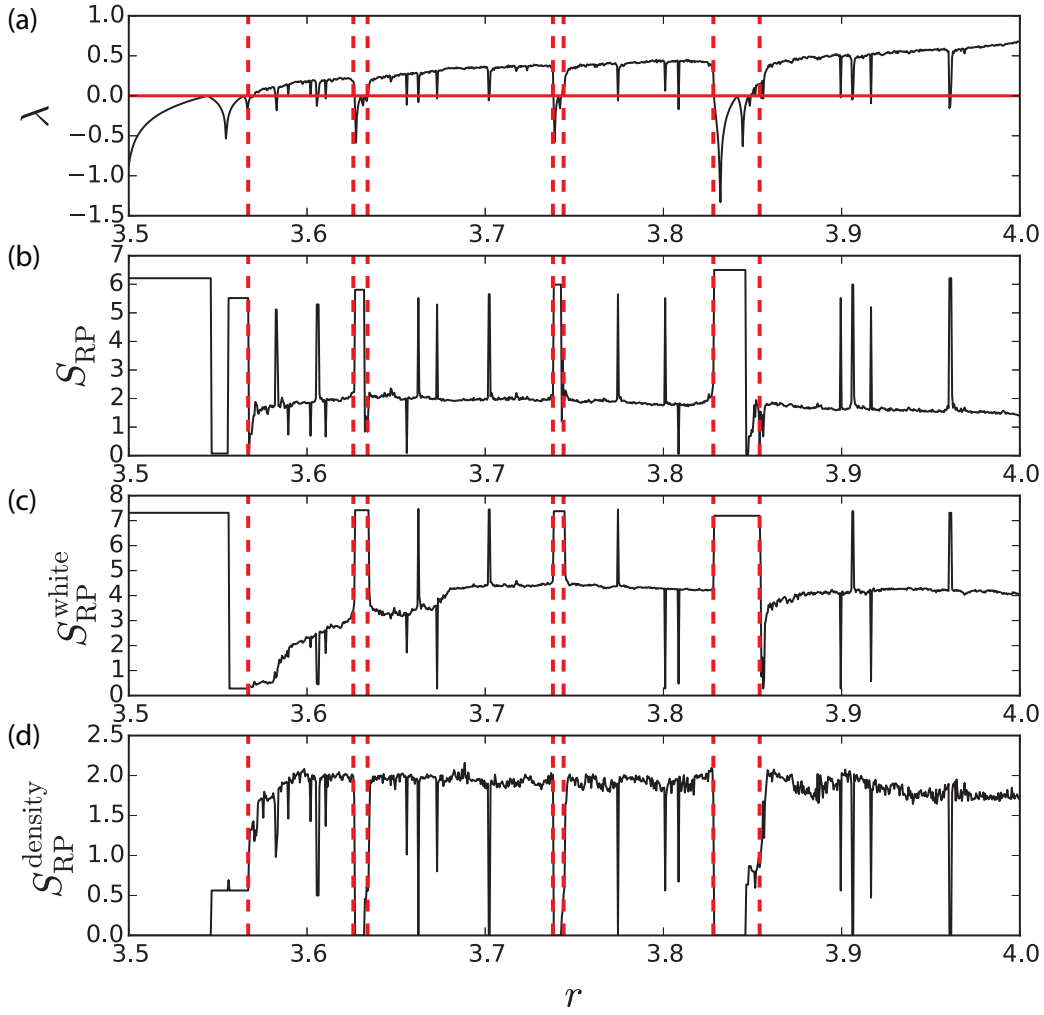

Figure 3.6: Comparison between (a) the Lyapunov exponent λ, (b) S_{RP} Eq. (3.3), (c) $S_{\mathrm{RP}}^{\mathrm{white}}$ Eq. (3.7) and (d) $S_{\mathrm{RP}}^{\mathrm{density}}$ Eq. (3.9) for the logistic map. Although the diagonal lines based entropies S_{RP} and $S_{\mathrm{RP}}^{\mathrm{white}}$ show anti-correlated behaviour with the Lyapunov exponent, density based entropy $S_{\mathrm{RP}}^{\mathrm{density}}$ is positively correlated. The red dashed lines show the critical transition points and all measures are sensible with them.

Obviously, the density based entropy $S_{\mathrm{RP}}^{\mathrm{density}}$ is more efficient measure to distinguish the behaviour of the system than the diagonal line based ones. Since its variation by the control parameter r is similar to the Lyapunov exponent. Moreover, the absolute Pearson correlation coefficient of the density based entropy $S_{\mathrm{RP}}^{\mathrm{density}}$ is much higher than the diagonal based ones S_{RP} and $S_{\mathrm{RP}}^{\mathrm{white}}$.

3.3 Entropy of Weighted Recurrence Plot

The selection of the arbitrary threshold distance ε to define the recurrences is one of the challenges often encounter in recurrence based techniques. As mentioned before, there are several ways to select threshold values for a time series (Mindlin and Gilmore, 1992; Koebbe and Mayer-Kress, 1992; Zbilut and Webber, 1992; Thiel et al., 2002; Schinkel et al., 2008), but there is no certain technique to choose the best threshold ε. In the case of density based recurrence measures such as $S_{\mathrm{RP}}^{\mathrm{density}}$, one can discard arbitrary threshold parameter. In this section, I will introduce an additional new recurrence based tool that comes without the need of choosing a threshold value called weighted recurrence plot (wRP) and will introduce the associated Shannon entropy.

3.3.1 Weighted Recurrence Plot

The restriction of regular RPs naturally introduce a free parameter in the analysis through the threshold distance. Instead of considering a binary matrix (regular RP), I take into account the normalized distance matrix $W_{i,j} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$. Note that the binary matrix $\mathbf{R}$ provides the information whether two points i and j are close or not in a d-dimensional phase space, whereas $\mathbf{W}$ represents the distances between point pairs of the time series (sometimes also referred to as unthresholded RP (Sipers et al., 2011)). In order to consider proximity between points of the time series, I introduce the weight matrix

$$\widetilde{W}_{i,j} = e^{-\|\mathbf{x}_i - \mathbf{x}_j\|} \tag{3.11}$$

which will be the base for our further analysis (Fig. 3.7). As particular weights $\widetilde{W}_{i,j}$ I have chosen the inverse of the exponential distribution, because it scales the distances to the value range [0, 1], with value one for close states and zero for non-close states. For example, periodic regimes with the occurrence of identical states at i and $j = 2\pi ki$ result in $\widetilde{W}_{i,j} = 1$, whereas for diverging states with large distances $\widetilde{W}_{i,j}$ will tend to zero. This scaling behaviour is correlated with regular RP and therefore we call this new tool *weighted recurrence plot*.

In the definition of weighted recurrence plot ($\widetilde{\mathbf{W}}$), there is an additional benefit, i.e, it does not require selection of the recurrence threshold value ε. Therefore weighted RPs overcome this limitation and are independent of the arbitrary choice of threshold ε.

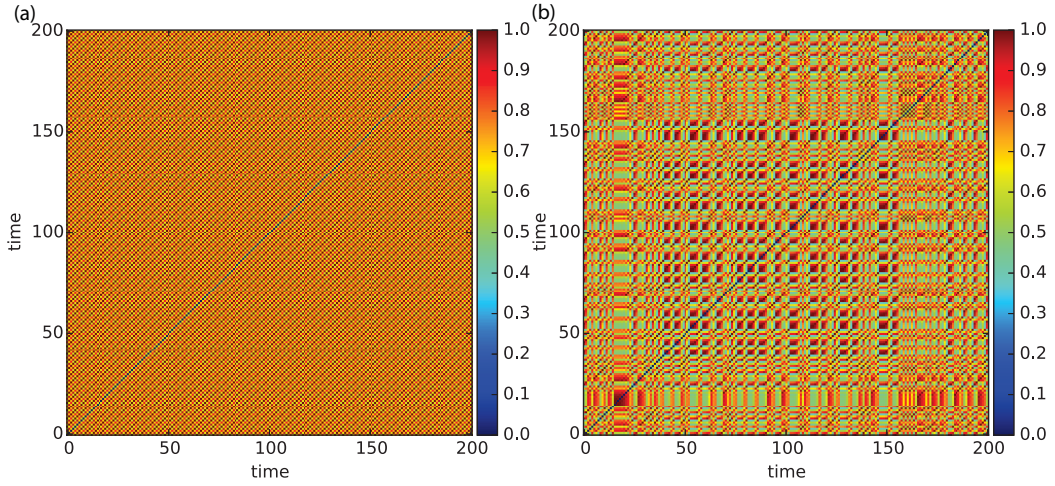

Figure 3.7: Weighted RPs for different nonlinear dynamics for 200 time steps: (a) periodic case for the logistic map, $x_{t+1} = rx_t(1 - x_t)$ system with $r = 3.5$, (b) chaotic case for the logistic map, $r = 4.0$.

Entropy of wRP

In order to define the Shannon entropy from wRP, I will use the same approach as I introduced previously in regular RP. First I define the strength s_i of a point i in the time series by

$$s_i = \sum_{j=1}^{N} \widetilde{W}_{ij}. \tag{3.12}$$

As in the case of RP, the strength s_i is the density heterogeneity at the time i and one can define the associated Shannon entropy as

$$S_{\mathrm{wRP}} = -\sum_{\{s\}} p(s) \log p(s), \tag{3.13}$$

where $p(s) = P(s)/S$ is the relative frequency distribution of the distance matrix strength and $S = \sum_i^N s_i$ is the total number of strengths.

Due to the binary recurrence matrix and the weighted distance matrix being positively correlated to each other (if two points i and j are identical $\widetilde{W}_{ij} = 1$ and $R_{ij} = 1$ or far away between each other $\widetilde{W}_{ij} \to 0$ and $R_{ij} = 0$), the expected outcomes of entropies $S_{\mathrm{RP}}^{\mathrm{density}}$ and S_{wRP} are positively correlated as well. Arbitrarily selected parameters are always a handicap for tools since the result of analysis strongly dependent on these parameters. It is quite important to find an invariant result for a particular time series via using only the system's properties. Therefore, the wRP approach is more consistent tool than the RP.

3.3.2 Applications

After I defined the new distance based tool and its associated entropy, I compare the results using the logistic map. Furthermore, I demonstrate the power of the technique on the continuous Rössler and on experimental electrochemical data.

Logistic Map

In order to compare the results of S_{wRP} with $S_{\mathrm{RP}}^{\mathrm{density}}$, first, I use time series data from the logistic map having the same properties as in Sec. 3.2.1. To evoke the technical details again: the map is investigated for the control parameter $r \in [3.5, 4.0]$ with a step size of $\Delta a = 0.0005$. For each value of r, I compute a time series of length $N = 5000$. In order to discard transients, I use only the last 3000 values of this time series.

Note that, in order to calculate S_{wRP}, 50 bins are used to derive the probability density function of strengths $p(s)$. As expected, S_{wRP} is also positively correlated with the Lyapunov exponent of the logistic map for the all range of the control parameter r as the density entropy of regular RP $S_{\mathrm{RP}}^{\mathrm{density}}$. Within periodic states, $S_{\mathrm{wRP}}(s)$ has lower values and during chaotic regimes, it has higher values (Fig. 3.8). At the critical values of $a = 3.544, 3.564, 3.630, 3.741$, and 3.841, $S_{\mathrm{wRP}}(s)$ reveals sharp jumps.

The correlation between $S_{\mathrm{wRP}}(s)$ and the Lyapunov exponent is estimated as,

$$\rho(S_{\mathrm{wRP}}(s), \lambda) = 0.85. \tag{3.14}$$

Positive correlation is found as expected. In the case of regular RP, the correlation between $S_{\mathrm{RP}}^{\mathrm{density}}$ and the Lyapunov exponent λ is positive as well but it is less than the correlation between $S_{\mathrm{wRP}}(s)$ and λ.

Rössler oscillator

The next application is a continuous flow, it is a mathematical model defined with coupled differential equations by Otto Rössler (Rössler, 1976),

$$\left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) = (-y - z, x + ay, b + zc), \tag{3.15}$$

where the bifurcation parameter is $b \in [0.0, 2.0]$, and the other parameters are $a = 0.2$ and $c = 5.7$. Here I analyse the Poincaré section of the y-component of the Rössler oscillator. For this particular control parameters set, the behaviour of the dynamics is contrary to the logistic map, increasing the control parameter b drives the system from the chaotic regime to the periodic ones. This behaviour can seen easily from the Lyapunov exponent Fig. 3.9(a). For each b value, I create a time series of the length $N = 6 \times 10^5$ and exclude the transient responses by using only the last 5×10^4 data points for the following analysis. I applied the Poincaré section on the phase space and use only the inner points of the Poincaré section to calculate $\widetilde{\mathbf{W}}$ and to estimate

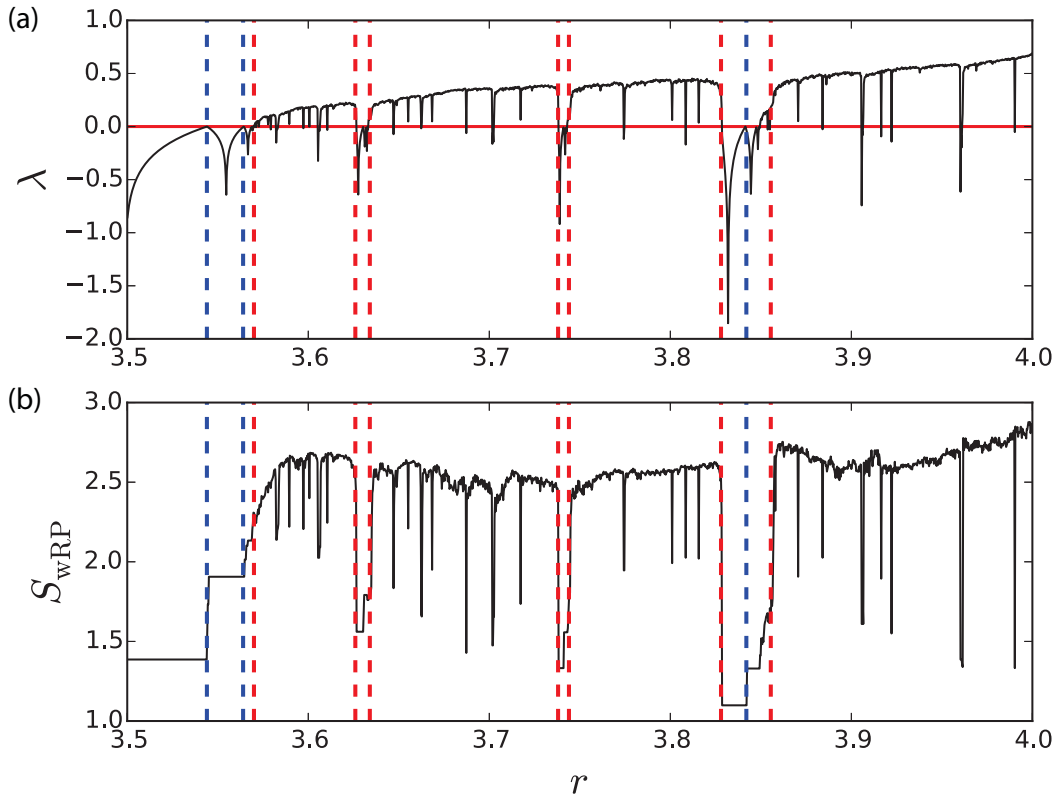

Figure 3.8: Comparison between (a) the Lyapunov exponent λ and (b) $S_{\mathrm{wRP}}(s)$ for the logistic map. The correlation between λ and $S_{\mathrm{wRP}}(s)$ is 0.85 and higher than other entropies S_{RP}, $S_{\mathrm{RP}}^{\mathrm{white}}$ and $S_{\mathrm{RP}}^{\mathrm{density}}$. Moreover, $S_{\mathrm{wRP}}(s)$ is not dependent on any arbitrary parameter such as recurrence threshold ε. Furthermore, while $S_{\mathrm{wRP}}(s)$ is perfectly sensible to periodic to chaotic transition points (red dashed lines) as well as it is possible to detect periodic to periodic transitions (blue dashed lines) with the same measure.

the Shannon entropy of wRP $S_{\mathrm{wRP}}(s)$. For each control parameter b, approximately $1,000$ inner points of the Poincaré section are used to compute the RPs. In order to calculate S_{wRP}, I also use 50 bins to construct the probability density function of strengths $p(s)$ as in the logistic map case.

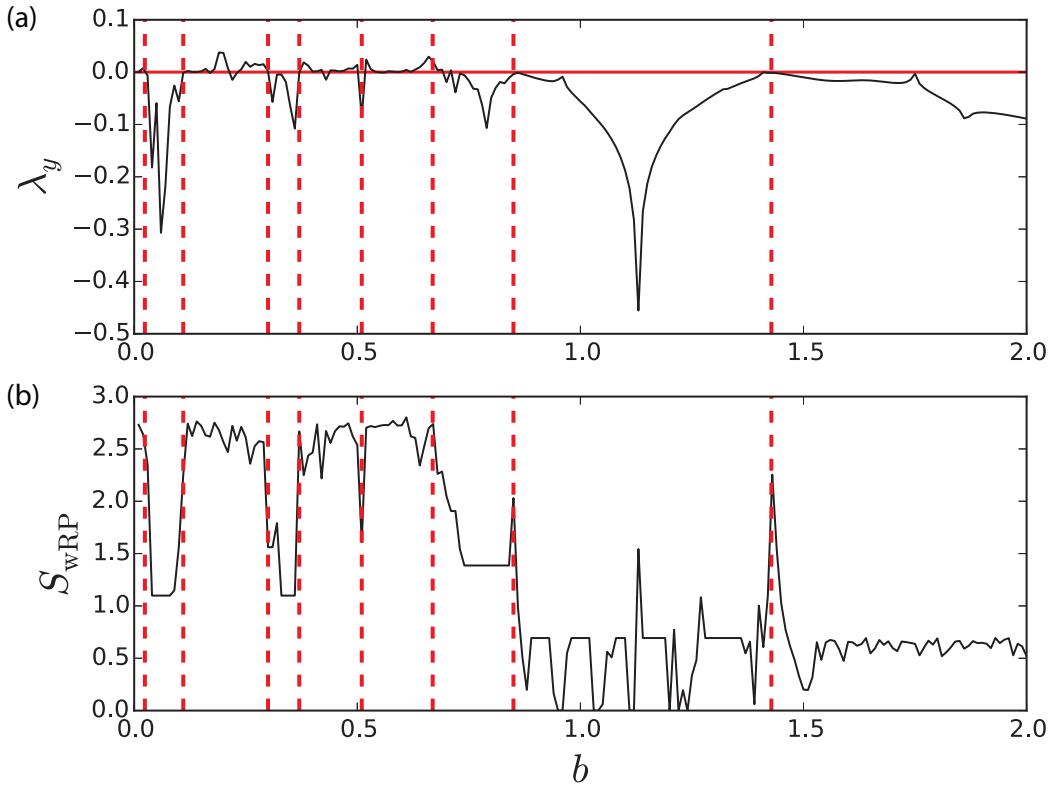

Figure 3.9: Comparison between (a) the Lyapunov exponent λ and (b) $S_{\mathrm{wRP}}(s)$ for the Rössler oscillator. The Poincaré section over the y-component of the Rössler oscillator is used for the analysis. Therefore, I compare the results with the Lyapunov exponent of y-component of the system, λ_y. The critical transition points are plotted by red dashed lines.

I estimate the wRP entropy S_{wRP} for the Rössler oscillations just like for the logistic map. As expected, the general tendency is similar to the findings of the logistic map example: during periodic regimes, S_{wRP} is lower than during the chaotic regimes (Fig. 3.9(b)). At the critical values of b, $S_{\mathrm{wRP}}(s)$ reveals abrupt changes.

Electrochemical Oscillators

The calculation of Lyapunov exponents is often infeasible for systems whose equations of motion are not accessible, thus various estimators for measuring the divergence behavior of nonlinear dynamics have been suggested (Kantz and Schreiber, 1997).

Now I will consider an application from electrochemistry, and apply the wRP approach on the experimental data sets. The goal is to demonstrate the capability of the wRP framework to grasp the complexity of patterns emerging from a real process.

The kinetics of nickel electrodissolution include charge transfer steps that exhibit negative differential resistance (NDR); this NDR behavior generates a range of nonlinear behavior in which the rate of metal dissolution (current) exhibits simple and complex periodic and chaotic behavior (Lev et al., 1989; Kiss et al., 2002; Kiss et al., 2002). In the experiment, 30 nickel wires (1mm diameter) are employed as working electrodes in an electrochemical cell with $Hg/Hg_2SO_4/(sat)K_2SO_4$ reference and a Pt coated Ti rod counter electrodes in 4.5 M sulfuric acid solution at 10^oC. The Ni wires were connected through 1.2 kΩ external resistances to a potentiostat (GillAC, ACM Instruments) that maintains constant circuit potential (V). The current time series for each wire (oscillator) are digitized at a rate of 200 Hz. Previous investigations showed that in this configuration there is negligible coupling among the wires and the oscillators can be regarded as independent (Koper, 1996). In addition, there exists an inherent heterogeneity due to varying surface conditions that creates a population of oscillators with slightly different dynamical characteristics, e.g., there is a frequency distribution with a range of about 10-20 mHz (Koper, 1996). The potential was incrementally increased by 10 mV from 1.3 to 1.4 V and about one thousand oscillations were collected at each potential. The frequency of the oscillations changed from 1.9 Hz (1.3 V) to 1.8 Hz (1.4 V). Transitions from simple periodic oscillations through complex oscillations to chaotic oscillations and back to complex oscillations was visually observed with increase of the potential. Inspection of the data reveals the waveform of oscillations were preeminently period-1 (1.30 to 1.33 V) (P_1), period-4 (1.34 to 1.35 V), chaos (1.36 to 1.39 V), and period-3 (1.4 V) (P_3). This data is consistent with previous observation that a complex behavior is generated with a period-doubling bifurcation to chaos with the presence of intermittent periodic windows (Lev et al., 1989; Kiss et al., 2000).

The time series from 30 electrode for different values of voltage are analyzed by S_{wRP}. The pattern of Fig. 3.10 shows that the periodic region (1.30-1.35) has small values of entropy and the chaotic region (1.36 to 1.39 V) has larger values. There are differences between the entropies of the individual oscillators since this is real experimental data. That means each oscillator can be affected from different noise, or oscillator can slightly differ from each other and so on.

In order to see the general outcome from all this oscillators, I averaged the results over all 30 oscillators Fig. 3.11. Because of the average, the trends in the changes of the regimes are better seen as the circuit potential is varied. It is easier to see the trend of the entropies due to changing the voltage of the system. The averaged entropy of wRP S_{wRP} is again correlated with the behaviour of the system.

Overall, I conclude that the S_{wRP} is suitable to reveal the complexity of the chemical reaction process; it has normalized values less than 0.2 for simple periodic oscillations, 0.2 to 0.7 for complex periodic (e.g., period 4 and period 3) oscillations, and larger than 0.7 for chaotic oscillations.

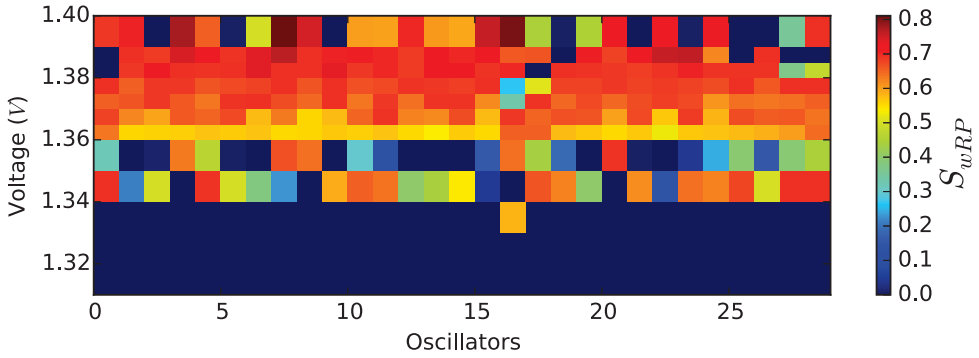

Figure 3.10: Entropy of wRP S_{wRP} of 30 electrochemical oscillators as a function of circuit voltage. Colormap shows the entropy of wRP associated to the time series of each electrodes for different voltages.

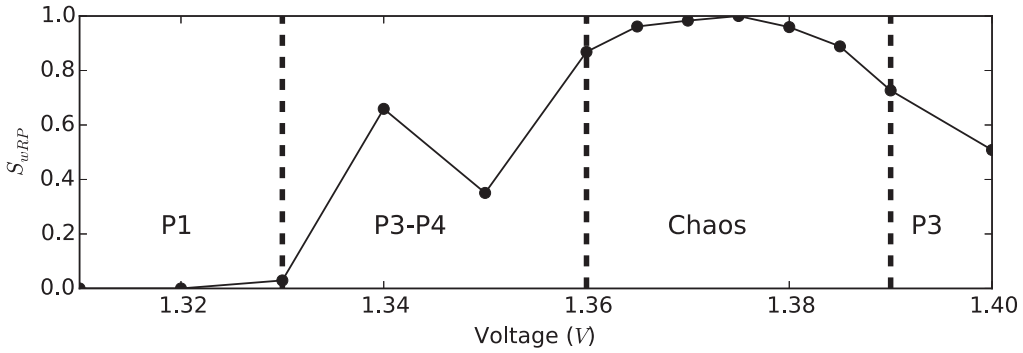

Figure 3.11: Averaged normalized entropy of wRP S_{wRP} over all 30 electrochemical oscillators for different voltages.

3.4 Summary

In this chapter I introduced the meaning of entropy and showed how entropy behaves as function of the frequency distributions. In the nature of RP, there are matrix limitations and due to these boundaries, probability distribution functions of diagonal lines are not fair. Therefore, when one use the Shannon entropy as a complexity measure, the outcome is anti-correlated with the Lyapunov exponent. In order to overcome this, I developed a new density based the Shannon entropy from regular RP, $S_{\mathrm{RP}}^{\mathrm{density}}$. I compared this new definition with pre-existing diagonal based entropies. I found that the density based measure is positively correlated and it is more correlated than absolute value of the diagonal based ones. However, there is a pitfall for this measure, in general, random processes are uniformly distributed in a phase space. Therefore, density based entropy definition cannot distinguish the behaviour either periodic or stochastic and is suggested for deterministic systems.

In order to get rid of the handicap of arbitrary threshold selection, I introduced the weighted recurrence plot. I applied the density based entropy definition on this new unthresholded approach and tested it on a discrete map (the logistic map), a continuous flow (the Rössler oscillator) and a real-world experiment (electrochemical oscillators). All results show that the wRP entropy S_{wRP} is a powerful measure and well suited to detect transitions in the underlaying dynamics. Specifically, where the Lyapunov exponent calculation is not possible due to limited number of time series points, of real data sets, S_{wRP} detects different regimes very significantly.

Chapter 4

Threshold Selection for Recurrence Networks

In nature and man-made systems, one of the most important structures are networks. Examples include collection of coupled biological and chemical systems, brain networks, power grid networks, superconductors, geological models, ecosystems etc. Naturally, it is very interesting to understand these structures in detail. In the last few decades, elaborative studies of real world complex systems depend on our understanding of the underlying structural features of networks. In order to analyze networks quantitively, many measures have been developed, based on structural characteristics.

During the last decades, network theory became a very popular tool to analyze time series as well. Network measures associated to local couplings and global structure have been used to determine the complexity or underlying dynamics of systems. In other words, networks are constructed from time series to investigate their structural properties. There are different ways to create networks from time series. In this chapter I will briefly mention these different network reconstruction techniques and focus on the recurrence based one, namely *recurrence networks* (RN). As in the recurrence plot case, RNs have an arbitrary threshold value too. Hence, the threshold parameter is very important in order to avoid bias of the analysis.

In this chapter, I propose a new method to select the recurrence threshold adaptively for a particular time series, the method was published in Nonlinear Processes in Geophysics (Eroglu et al., 2014a). I compare the results with constant and adaptive thresholds to study periodic–chaos and even periodic–periodic transitions in the dynamics of a prototypical model system. This novel method is then used to identify climate transitions from a lake sediment record.

4.1 Graph Theory

Graph theory is the mathematical framework to investigate networks. Graphs are formed by a set of nodes (vertices) and a set of edges (links) that connect the nodes (Fig. 4.1). Each network is a graph and can be investigated by graph theory.

As mentioned before, there are many kinds of different networks in nature and in man-made systems. In order to understand different structures, these many network models have been developed. If there are no self-loops for any nodes, a network is called *simple* network (Fig. 4.1). An *undirected* graph where all edges are bidirectional,

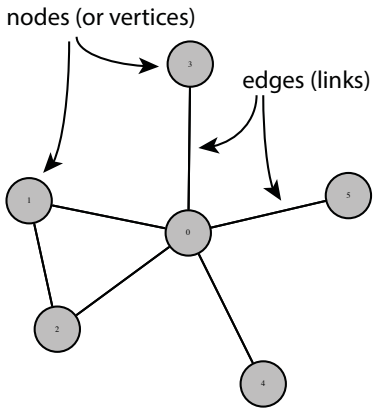

Figure 4.1: A simple graph.

means if node-i is connected to node-j with a bidirectional link, node-i is in interaction with node-j and node-j is in interaction with node-i (Fig. 4.2(a)). Undirected simple networks are not always very realistic. For instance, World Wide Web is one of the most famous network and there are links between web pages. These links are one way directed ones, and the network is called *directed* (Fig. 4.2(c)). The creation of more complex networks is possible and different kind of nodes or weighted edges and so on can be used. But in this thesis I focus on undirected and unweighted networks. More details can be found in (Newman, 2010).

A *path* in a graph is a sequence of connected nodes, without repetition. The *diameter* of a network d determines the greatest length of the shortest path between any two nodes on the graph. In order to find the diameter of a graph, first find all shortest path length and the greatest one is the diameter of the network. If there is an isolated part in the graph, one or more nodes are unconnected with the network, then the diameter of the network is infinite. A network with finite diameter is called *connected* (Fig. 4.2(a)), otherwise *unconnected* (Fig. 4.2(b)).

4.1.1 Adjacency and Laplacian Matrices

In graph theory, connections between nodes can be described with the adjacency matrix $\boldsymbol{A}$, which holds the topological information of the graph structure, i.e.,

$$A_{i,j} = \begin{cases} 1, & \text{if node-}i \text{ and node-}j \text{ are connected} \\ 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

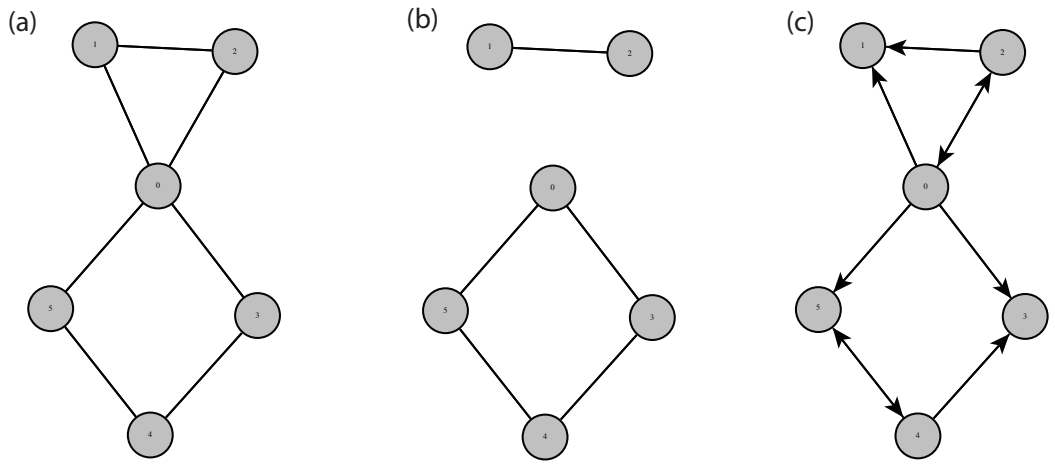

Figure 4.2: Various graph examples: (a) connected graph, (b) unconnected graph, and (c) directed graph

The adjacency matrix of an undirected network is a symmetric matrix. The *degree* k_i is total number of edges connected to the ith node in the graph. It is easy to describe it mathematically by,

$$k_i = \sum_{j=i}^{N} A_{i,j} \tag{4.2}$$

The Laplacian matrix $\boldsymbol{L}$ is another way to represent a network by,

$$L_{i,j} = \begin{cases} k_i, & \text{if } i = j \\ -1, & \text{if node-}i \text{ and node-}j \text{ are connected} \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

There is a direct relationship between the Laplacian and the adjacency matrix. The relationship is given by,

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}, \tag{4.4}$$

where $\boldsymbol{D} = diag(k_1, k_2, ..., k_n)$ is the diagonal matrix of degrees. The Laplacian matrix will be the backbone of my investigation in the following chapters. I will use the eigenvalue spectrum of the Laplacian matrix to find an optimal parameter for reconstructing a network from time series.

The spectrum of the Laplacian matrix $\boldsymbol{L}$ holds a lot of structural information of networks such as algebraic connectivity density of a graph or spreading speed of a random walk on a graph etc. In my study, the most important eigenvalue of the Laplacian matrix is the smallest nonzero one (λ_2) since it tells us if a network is

connected or not.

Now I consider G is an undirected networks and the associated Laplacian matrix is $\boldsymbol{L}$. Clearly, L is a symmetric matrix and hence its eigenvalues are real. To proof this I can simply write,

$$\boldsymbol{L} = \boldsymbol{L}^{\dagger} \tag{4.5}$$

where $\boldsymbol{L}^{\dagger}$ is the Hermitian conjugate of $\boldsymbol{L}$.

$$\boldsymbol{L\nu} = \lambda\boldsymbol{\nu}, \tag{4.6}$$

where ν is the eigenvector and λ is the eigenvalue of $\boldsymbol{L}$. If I left-multiply with an eigenvector ($\nu^{\dagger}$) to Eq. (4.6),

$$\boldsymbol{\nu}^{\dagger}\boldsymbol{L\nu} = \boldsymbol{\nu}^{\dagger}\lambda\boldsymbol{\nu} = \lambda\boldsymbol{\nu}^{\dagger}\boldsymbol{\nu} \tag{4.7}$$

$$(\boldsymbol{\nu}^{\dagger}\boldsymbol{L\nu})^{\dagger} = (\lambda\boldsymbol{\nu}^{\dagger}\boldsymbol{\nu})^{\dagger} \tag{4.8}$$

$$\boldsymbol{\nu}^{\dagger}\boldsymbol{L\nu} = \lambda^{\dagger}\boldsymbol{\nu}^{\dagger}\boldsymbol{\nu} \tag{4.9}$$

$$\lambda = \lambda^{\dagger} \tag{4.10}$$

where $(\boldsymbol{\nu}^{\dagger}\boldsymbol{L\nu})^{\dagger} = \boldsymbol{\nu}^{\dagger}\boldsymbol{L\nu}$. It follows that Hermitian matrices have real eigenvalues.

The eigenvalues of the Laplacian matrix $\boldsymbol{L}$ cannot be negative. Assume that, the network G has n nodes and m edges. Now if I consider each edge's ending points tagged as first and second end point randomly. It is of no importance how one selects which end is the first and which one is second. Therefore, one can determine a matrix $\boldsymbol{B}$ with the size of $m \times n$,

$$B_{i,j} = \begin{cases} 1, & \text{if edge-}i\text{ connected to node-}j\text{ with first end point} \\ -1, & \text{if edge-}i\text{ connected to node-}j\text{ with second end point} \\ 0, & \text{otherwise.} \end{cases} \tag{4.11}$$

This algorithm gives us such a matrix $\boldsymbol{B}$ that one element is 1, another one element is -1 and the rest zeros for each row in the matrix. There is a relationship between $\boldsymbol{B}$ and the Laplacian matrix $\boldsymbol{L}$. Multiplication of the elements of matrix $\boldsymbol{B}$ gives $\boldsymbol{L}$, $L_{i,j} = \sum_k B_{k,i}B_{k,j}$. According the Cholesky decomposition, I can write it in matrix multiplication form as,

$$\boldsymbol{L} = \boldsymbol{B}^T\boldsymbol{B}, \tag{4.12}$$

where $\boldsymbol{B}^T$ is the transpose of $\boldsymbol{B}$. If λ_i is the eigenvalue of ν_i for the Laplacian matrix $\boldsymbol{L}$, one can write the equation below,

$$\boldsymbol{\nu}_i^T\boldsymbol{B}^T\boldsymbol{B\nu}_i = \boldsymbol{\nu}_i^T\boldsymbol{L\nu}_i = \lambda_i\boldsymbol{\nu}_i^T\boldsymbol{\nu}_i = \lambda_i, \tag{4.13}$$

Obviously, λ_i are given by the scalar multiplication of $(\boldsymbol{B\nu}_i)^T$ and $\boldsymbol{B\nu}_i$. Therefore, λ_i is derived from the summation of squared real numbers and cannot be negative.

$$\lambda_i \geq 0 \quad \text{for all } i = 1, 2, ..., n \tag{4.14}$$

Since the row sum is zero, the Laplacian matrix has one zero eigenvalue with an associated eigenvector of ones $\mathbf{1}$. It is easy to proof that by multiplying the Laplacian matrix $\boldsymbol{L}$ with $\mathbf{1} = (1, ..., 1)^T$,

$$\sum_j L_{i,j} \times \mathbf{1} = \sum_j (\delta_{i,j} k_i - A_{i,j}) = k_i - \sum_j A_{i,j} = k_i - k_i = 0 \tag{4.15}$$

One can simply summarize the properties of eigenvalues of the Laplacian until now by,

$$0 = \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n, \tag{4.16}$$

the equation follows the Gershgorin theorem (Varga, 2010).

The number of zero eigenvalues denotes to the number of unconnected components in the network. Assume that a network G consists of r connected subcomponents $G_1, G_2, ..., G_r$ and each subcomponent has its own Laplacian matrix $\boldsymbol{L}_1, \boldsymbol{L}_2, ..., \boldsymbol{L}_r$. Then, the Laplacian $\boldsymbol{L}$ of the giant network G splits into the Laplacian of subcomponents $\boldsymbol{L}_1, \boldsymbol{L}_2, ..., \boldsymbol{L}_r$.

Let m be number of zero eigenvalues. Then each $\boldsymbol{L}_i$ has an eigenvector $\boldsymbol{\nu}_i$ with zero eigenvalue. The eigenvector $\boldsymbol{z}_i = (z_i^1, ...z_i^n)$ of $\boldsymbol{L}$ can be defined as z_i^j equals 1 if j is an element of the ith subgraph and zero otherwise, thus $m \geq r$. In order to complete the proof, I need to show that entries of any eigenvector $\boldsymbol{g}$ associated with zero eigenvalue are constant. I assume that g is a non constant eigenvector associated with zero eigenvalue, and the largest element of the eigenvector is $g_\ell > 0$. Under these assumptions, it is clear that $(\boldsymbol{Lg})_\ell = 0$ since the eigenvalue is zero.

$$(\boldsymbol{Lg})_\ell = \sum_j L_{\ell,j} g_j = \sum_j (k_\ell \delta_{\ell j} - A_{\ell,j}) g_j = 0 \tag{4.17}$$

then g_ℓ equals to the mean of the values of $\boldsymbol{g}$ assigned to its neighbours by,

$$g_\ell = \frac{\sum_j A_{\ell,j} g_j}{k_\ell} \tag{4.18}$$

Therefore, all elements of $\boldsymbol{g}$ are equal, in other words, $\boldsymbol{g}$ must be a constant eigenvector. Moreover, it shows that the spectral gap λ_2 of the connected graph must be larger than zero.

The spectrum of the Laplacian matrix of a network contains more information about networks but I will use this particular feature in my novel work in the next section.

4.2 From Time Series to Networks

Network-based approaches have taken an important place in dynamical system analysis. There are several algorithms to generate networks from time series. The main goal of the transformations is to analyze and to quantify the feature of the time series. Complex network measurements are useful to investigate and understand the complex behavior of real world systems such as social, computer (Newman, 2002), or brain networks (Singer, 1999). The adjacency matrix of a complex network contains the structure of the system, thus, determines the links between the nodes of a network.

There are several ways to produce a network from time series data. The first algorithm is proposed by Zhang and Small in 2006 (Zhang and Small, 2006), and nodes were chosen from cycles in a time series. In order to link the nodes, they used a similarity measures. If cycle-i and cycle-j are sufficiently similar, node-i is connected to node-j. The advantage of this approach is that it is very robust to noise, but to construct a network, time series should be sufficiently oscillatory. Otherwise, there is no way to define cycles. In general, cycles of a time series are not enough to apply this technique. Therefore, the approach is improved by performing time delay phase space reconstructions (Xu et al., 2008). A time delay embedding allows to create more functional time series and analyze it in more details.

Another way to create a network from a time series is recurrence networks (RNs). The creation of RNs is very similar to recurrence plots (see Chap. 2), closeness of points in the phase space is the main idea to generate the network. RNs will be the main focus for the remaining part of this chapter.

Lacasa et al. (2008) introduced a new method, called *visibility graph*, based on correlation and self-similarities in the data. The visibility graph approach has been adopted from the field of computational geometry and has been introduced to detect mutual visibility relationships between points of a time series (Lozano-Pérez and Wesley, 1979; Nagy, 1994; Floriani et al., 1994; Turner et al., 2001). Thereafter visibility graphs have been applied to different fields such as geoscience and financial data (Donner and Donges, 2012). The main advantage of this method is that it is a well-suited tool to characterize different stochastic behaviour and self-similar sequences.

The last member of the family of network generator from time series has been proposed by Small (2013). This new approach depends on dynamical properties of a particular time series, instead of using closeness of nodes in the phase space or self-similarities of sequences. Nodes are created by dynamical pattern orders and linked based on temporal succession. In order to create this kind of network, differently from other phase space based techniques, time delay reconstruction is not necessary. Symbolic encoding of w points creates the nodes. Although using the dynamical properties of a time series is a very suitable approach to understand the underlying dynamics, the pitfall of this technique is the arbitrarily selected window size w.

4.2.1 Recurrence Networks

Recurrence networks are based on the recurrence matrix (Eq. (2.10)) which is a $N \times N$ matrix where N is the length of the phase space trajectory (the number of time steps). I now consider these time steps as nodes of a network; if the nodes are sufficiently close to each other, in other words, if the space vectors are neighbours, there is a link between them. In network theory, connections between network nodes can be described with the adjacency matrix $\mathbf{A}$, with $A_{i,j} = 1$ if there is a link between nodes i and j, otherwise $A_{i,j} = 0$. To obtain the adjacency matrix from the recurrence matrix, I discard self-loops in the recurrence matrix, i.e.,

$$A_{i,j} = R_{i,j} - \delta_{i,j}, \tag{4.19}$$

where $\delta_{i,j}$ is the Kronecker delta ($\delta_{i,j} = 1$ if $i = j$, otherwise $\delta_{i,j} = 0$).

As mentioned before, the number of links of the ith node (the degree) is given by $k_i = \sum_j A_{ij}$. In this work I use the eigenvalue spectrum of the Laplacian matrix $\mathbf{L}$ to find an adaptive threshold ε_c, where $L_{i,j} = \delta_{i,j} k_i - A_{i,j}$.

The crucial point in this study is choosing the adaptive closeness threshold (ε) for calculating the RN. A threshold for recurrence based methods should be *sufficiently* small (Marwan et al., 2007; Donner et al., 2010; Donges et al., 2012). Too small ε cause very sparsely connected RN with many isolated components; too large ε results in an almost completely connected network. For data sets which are not smooth, choosing an actually reasonable small threshold could nevertheless result in unconnected recurrence network components. These unconnected components would cause problems for some complex network measures, since some of them need a connected network to be computed for the entire network. For example, even if we have just one node that is not connected to the network, the average path length will be always infinite for the entire network. An even more important motivation for avoiding isolated components in the RN is that the RN provides a large amount of information about the dynamics of the underlying system although it contains only binary information. This has been demonstrated by reconstructing time series from RPs (Thiel et al., 2004b; Hirata et al., 2008). The condition for reconstructing a time series from a RP is that all points are connected by their neighborhoods, i.e., there are no isolated components. By applying recurrence measures I would like to quantify the dynamics encoded by the RN. This can be ensured by the above mentioned condition.

To find a *sufficiently* small threshold ε that fulfills the desired condition of connected neighborhoods, I will use the connectivity properties of the network. In particular, I choose the value for ε that is the smallest one for the RN to be connected. In order to find such an adaptive threshold, I start from very small values of the threshold and vary the ε parameter until I get a connected network. In order to apply this approach efficiently, I use an iterative bisection method in the simulations (Fig. 4.3).

As I have mentioned in Sec. 4.1.1, the connectivity of a network can be measured by the second smallest eigenvalue λ_2 of the Laplacian matrix, i.e. the network is

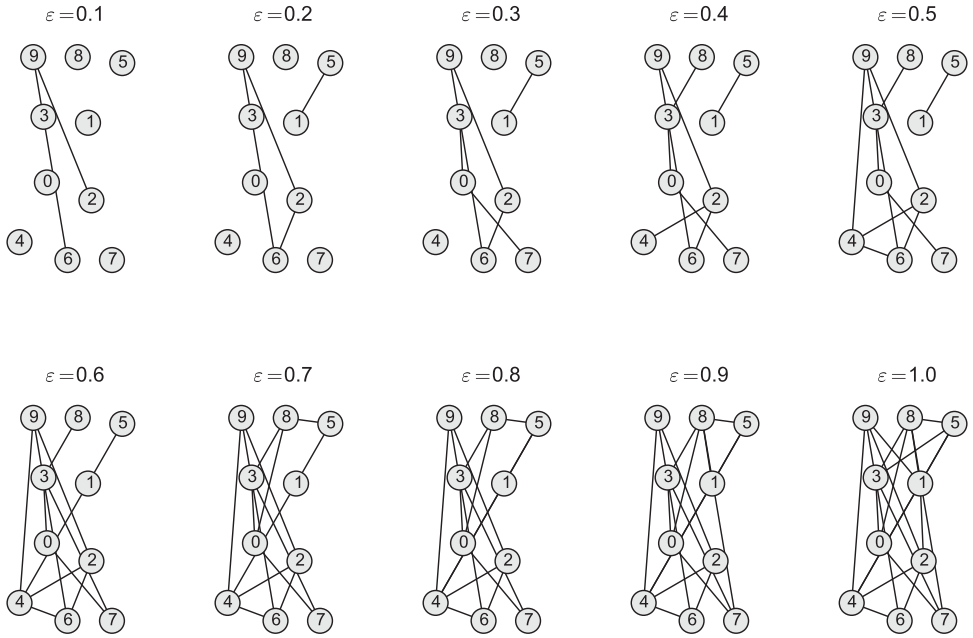

Figure 4.3: Variation of the threshold ε for very small segment of the logistic map as an illustrative example (control parameter $a = 4.0$). For a small ε value, network is not connected means some nodes are not involved in the structure such as $\varepsilon = 0.1, .., 0.5$. Therefore, when one analyzes these structures, there is no meaning of unconnected nodes. In other words, some points of a time series are ignored. Thus, a lot of information of time series is lost. For large ε values, there is no unconnected component anymore. According to the definition of threshold selection, it needs to be *sufficiently* small. Therefore, the best selection of a threshold is that connected network with the smallest ε. For this particular example, the threshold can be $\varepsilon = 0.6$.

connected $\lambda_2 > 0$ (Boccaletti et al., 2006). I choose the adaptive threshold value as the minimum value of the sequence of thresholds $\mathbf{T} = T_i, T_{i+1}, \ldots$ when the second minimum eigenvalue λ_2 is positive,

$$\varepsilon_c = \min(\mathbf{T}) \text{ with } \mathbf{T} = \{T_i \,|\, \forall i : \lambda_2(T_i) > 0\}. \tag{4.20}$$

Values ε below the critical value ε_c are indicating the existence of unconnected components in the RN (Fig. 4.4). After that critical threshold, λ_2 becomes positive and if I still increase the threshold the connectivity of the RN is increasing. By choosing the critical point ε_c as the recurrence threshold, I ensure that the RN will be connected by the smallest threshold possible.

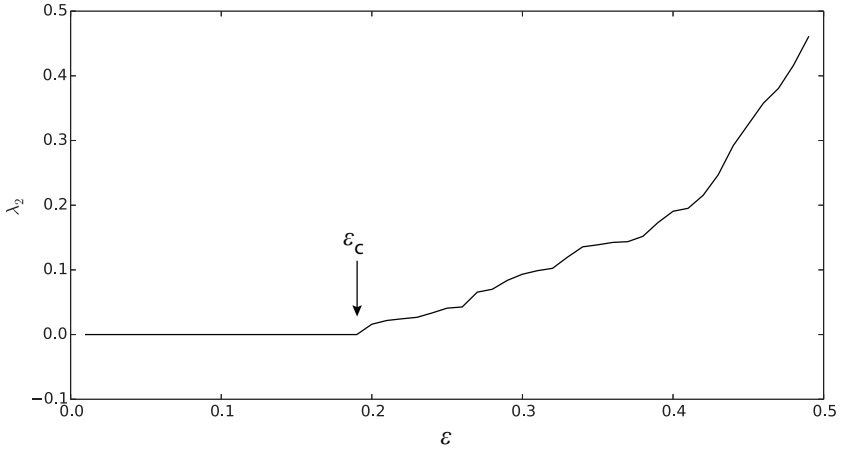

Figure 4.4: Variation of the second smallest eigenvalue of the Laplacian λ_2 due to changing threshold value, using the logistic map as an illustrative example (control parameter $a = 4.0$). $\lambda_2 = 0$ for thresholds below a critical value ε_c, indicating the existence of unconnected components in the RN. For $\varepsilon > \varepsilon_c$, there are no unconnected components in the RN anymore. The adaptive threshold value for this time series is $\varepsilon_c \approx 0.19$.

4.3 Applications

4.3.1 Logistic map

As a first application I compare some RN measures for using first the adaptive and then constant threshold approach by analysing the logistic map

$$x_{i+1} = a x_i (1 - x_i). \tag{4.21}$$

The detection of transitions between these different regimes was studied with RP and RN previously(Trulla et al., 1996; Marwan et al., 2009). I again focus on the range

$a \in [3.5, 4.0]$ with step size of $\Delta a = 0.0005$ and discard transients, I delete the first 2000 points of the time series of $N = 5000$ to avoid transients.

For the constant threshold selection method, I use the recurrence rate method to choose a threshold value: a threshold is selected in such a way that the recurrence rate RR is constant even for different time series with different dynamics (e.g., different values of a) (Marwan et al., 2007). In this work, I arbitrarily select $RR = 5\%$.

Now I compute the RNs by using the given threshold selection techniques ε and ε_c for each control parameter a. I then calculate transitivity T and betweenness centrality BC in order to detect the transitions from periodic to chaotic, chaotic to periodic states, bifurcations and inner(outer)-crisis. The network transitivity is given by,

$$T = \frac{\sum_{i,j,k} A_{i,j} A_{j,k} A_{k,i}}{\sum_{i,j,k} A_{k,i} A_{k,j}}, \quad T \in [0, 1]. \tag{4.22}$$

The average betweenness centrality of network,

$$BC = \frac{1}{N} \sum_{v} \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}, \quad BC \in [0, \infty[, \tag{4.23}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is number of those paths that pass through v. As mentioned in the previous section, not all complex network measures can be applied to a disconnected network. Therefore, it would cause problems to compute the measures on RNs resulting from the constant threshold technique, since the network could be disconnected. For instance, to compute the average shortest path length or assortativity for an entire network, the network must be connected (Newman, 2002; Newman, 2003). Disconnected nodes of the network could be discarded from the calculation, but in this case, I would lose information. In the adaptive threshold case, I could calculate all these measurements on the entire network since the selection of the adaptive threshold ensures that the recurrence network is connected.

Both threshold selection methods could detect transitions between dynamical regimes (periodic-chaos or chaos-periodic). Transitivity gives large values for the chaotic regime and small values for periodic. The betweenness centrality case is contrary to transitivity, and shows large values for periodic and small values for the chaotic regimes. Although the constant threshold selection detects the periodic windows (chaos-period transitions) more sharply than the adaptive threshold case, the transitivity and betweenness centrality for the constant threshold selection case, T_{constant} and BC_{constant}, cannot distinguish between different periodic dynamics, i.e., cannot detect certain bifurcation points such as for period doublings, e.g., at $a \approx 3.544, 3.564, 3.84$. Contrary, in the adaptively chosen threshold case, T_{adaptive} and BC_{adaptive} are sensitive to these bifurcations (Figs. 4.5, 4.6). Thus, using the adaptive threshold allows also the detection of periodic–periodic transitions (i.e., the study of bifurcation points where the maximal Lyapunov exponent stays non-positive).

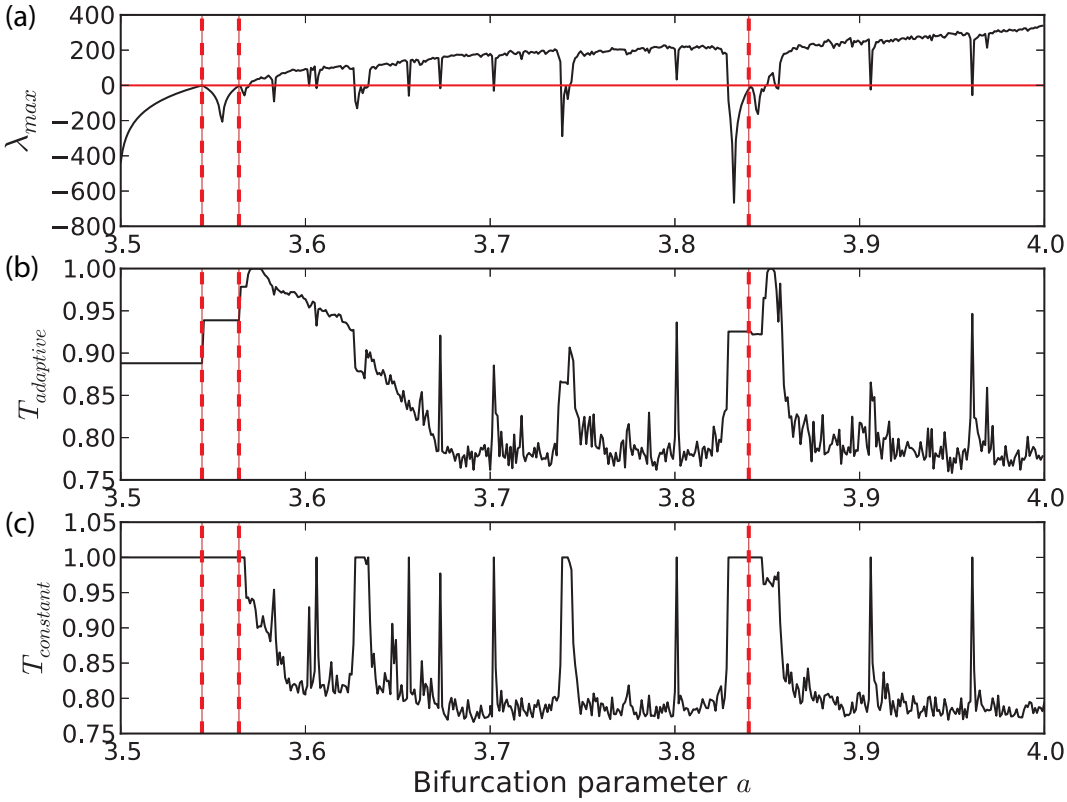

Figure 4.5: (a) Lyapunov exponent and transitivity using (b) adaptive threshold and (c) constant threshold for the logistic map. Dashed lines show certain bifurcation points before the chaotic regime.

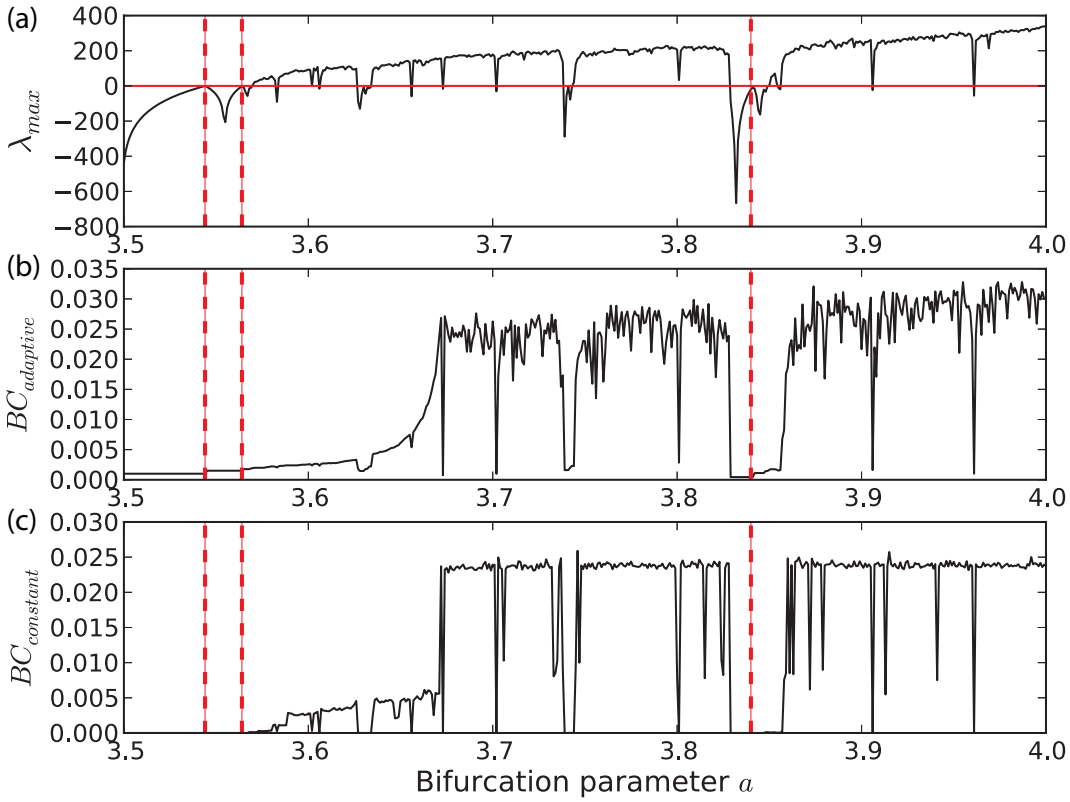

Figure 4.6: (a) Lyapunov exponent and betweenness centrality using (b) adaptive threshold and (c) constant threshold for the logistic map. Dashed lines show certain bifurcation points before the chaotic regime.

4.3.2 Application to palaeoclimate record

Lake sediments provide valuable archives of past climate variations. The study of palaeoclimate variation helps in understanding and evaluating possible future climate change.

In the following I will focus on a well dated high resolution climate archive from palaeolake Lisan located beneath the archaeological site of Massada in the Near East (Prasad et al., 2004; Prasad et al., 2009). The sediments from the Upper Member were deposited ($26 - 18$ cal ka BP) when the lake reached its highest stands (Bartov et al., 2003; Torfstein et al., 2013). The sedimentary sequence contains varves comprising seasonally deposited primary (evaporitic) aragonite and silty detritus (Prasad et al., 2004). The pure aragonite sublaminae were precipitated from the upper layer of the lake during summer evaporation. Their formation requires inflow of HCO_3^- ions into the lake from the catchment area during winter floods (Stein et al., 2003) that also bring in silty detrital material. One detrital and overlying aragonite sublaminae constitute a varve. Previous studies (Prasad et al., 2004; Torfstein et al., 2013) indicate that small ice-rafting events (denoted as a, b, c, and d), as well as prominent Heinrich events in the North Atlantic, are associated with the Eastern Mediterranean arid intervals. The study of seasonal sublaminae yields evidence of decadal to century scale arid events that correlate with cooler temperatures at higher latitudes. Analyses in the frequency domain indicate the presence of periodicities centered at 1500 yr, 500 yr, 192 yr, 139 yr, 90 yr, and $50 - 60$ yr, suggesting a solar forcing on climate (Prasad et al., 2004).

I use the yearly sampled pure aragonite proxy ($CaCO_3$) from the palaeolake Lisan for my RN analysis (Fig. 4.8a). I use a time delay embedding with dimension $m = 3$ and delay $\tau = 2$ (these parameters have been computed by standard procedure using false nearest neighbours and mutual information (Packard et al., 1980; Kantz and Schreiber, 1997)) for reconstructing the phase space. To detect dynamical transitions in the palaeoclimate data, I adopt a sliding window of W data points with a step size of ΔW. RNs are computed for each window of the time series one by one. I have chosen a sampling window size of $\Delta T = 100$ yr with 90% overlap corresponding to a time window size of $W \approx 100$ data points (since there are some gaps in the data, it is not exactly 100). The time series' length is $N = 7665$ and the total number of the windows analysed is

$$\frac{N - W}{\Delta W} \approx 755.$$

Transitivity and betweenness centrality is then calculated within these windows (Fig. 4.8b and c). As I have shown for the logistic map, transitivity and betweenness centrality are both sensitive to detect transitions. Larger values of transitivity T refer to regular behaviour, whereas smaller values to more irregular dynamics in the particular time series window .

The grey shaded horizontal band in Fig. 4.8b, c is the confidence interval of the network measures. I apply a simple test in order to see whether the characteristics of the dynamics at a certain time statistically differs from the general characteristics of

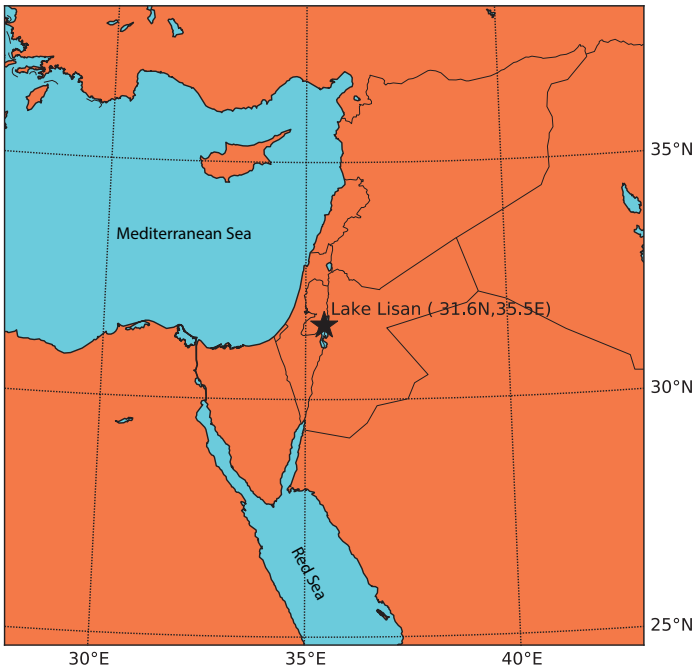

Figure 4.7: Map indicating the location of Lake Lisan: (31.6N, 35.5E).

the dynamics. In order to apply this statistical test, I use the following approach. I create surrogate data segments of length W by drawing data points randomly from the entire time series and I compute the RN and the network measures from such a surrogate segment. I repeat this 10,000 times and have an empirical test distribution of transitivity T and betweenness centrality BC. A confidence interval is then estimated from these distributions by their 0.05 and 0.95 quantiles.

Previous studies (Prasad et al., 2004) had identified multiple climate fluctuations in the varved Lisan record and correlated them with the Greenland oxygen isotope data (indicative of temperature changes, (Stuiver and Grootes, 2000)) and ice rafting events in the north Atlantic (Bond et al., 1997). The blue and orange vertical bars in Fig. 4.8 delineate periods of cooling and warming respectively in the higher latitudes that resulted in drier and wetter episodes in the eastern Mediterranean.

The network measures T and BC both indicate abrupt transitions (Fig. 4.8b, c). In particular for T, the values jump between high and low values. T reveals epochs of significantly low values at around 25.8–25.6, 25.2–25.1, 24.3–24.2, 24.0–23.9, 22.8–22.6, 22.3–22.1, 21.5–21.1, 21.7, 20.6–20.5, 20.1–19.9, 19.8–19.6, and 19.3–18.9 cal ka BP. The periods 25.8–25.6, 22.3–22.1, 21.5–21.1, and 19.3–18.9 cal ka BP correspond to the known Bond events d, c, b, and a, and the epoch between 24.3 and 23.9 cal ka BP coincides with the Heinrich H2 event. During the interstadial peaks IS2 at 23.8–23.7 and 23.3–23.2 cal ka BP, T shows significant high values, almost reaching the value

one. BC exhibits a rather similar behavior of abrupt transitions like T, but with opposite sign. A general observation is that low values in T can be found during dry but high values during wet regimes, and that such regimes change abruptly.

A high transitivity value indicates a more regular deposition of aragonite, and, thus, a more regular, or even periodic climate variability. This could be an indication for a dominant role of the (more or less periodic) solar forcing via its influence on the temperature in the higher latitudes. During phases of a colder North Atlantic, the solar forcing become less important but regional climate effects more important and therefore dominating, causing a more complex, irregular climate variability, finally indicated by low values of T.

Combining the maxima of T and minima of BC, I can identify the above mentioned periods of non-regular climate dynamics. Most of these periods correspond to cold events, e.g., the Bond, Heinrich events as well as the Lisan lake events L3 till L13 (Prasad et al., 2004). Several regular periods can be identified, some of them coinciding with the warm period during the interstadial IS2. Few remaining periods of high or low regularity have not yet been identified in the literature so far and call for further investigation.

The abrupt changes in T are available due to the adaptive threshold. By using a constant threshold, T varies only slowly and more gradual. Identifying the accurate times of climate regime shifts becomes more difficult in this case.

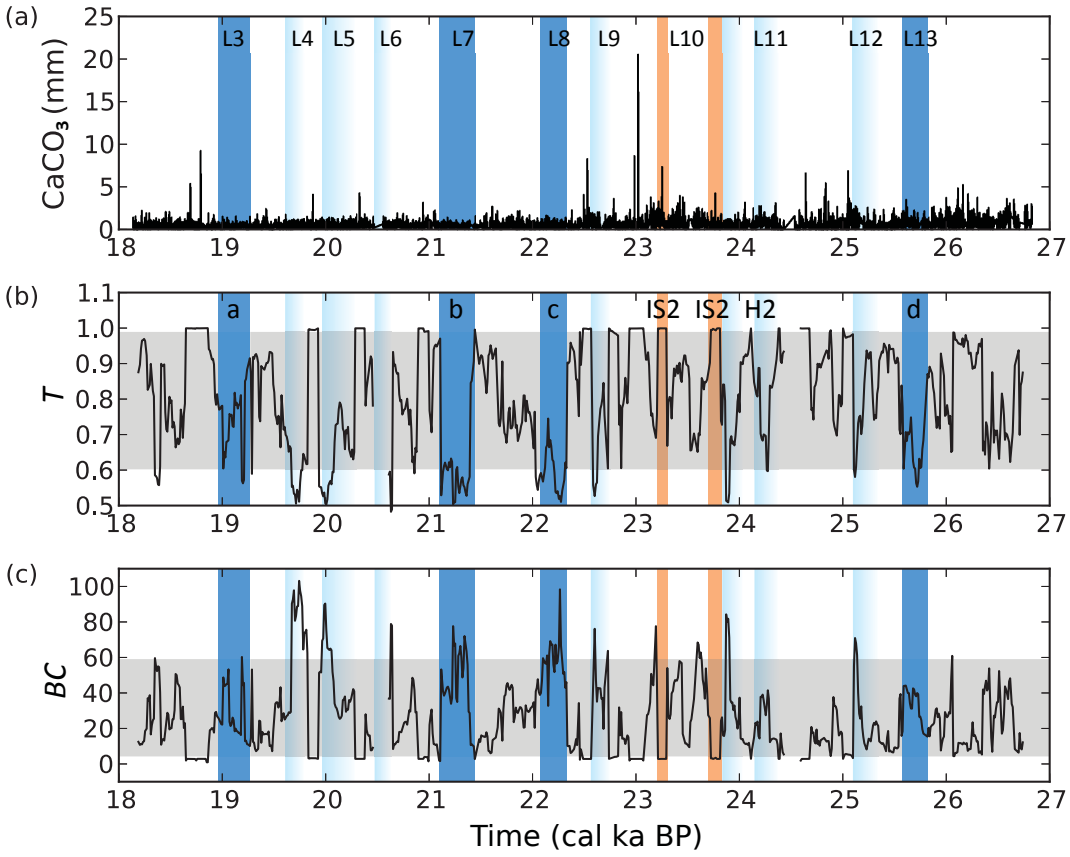

Figure 4.8: (a) Aragonite ($CaCO_3$) record from palaeolake Lisan, (b) transitivity, and (c) betweenness centrality results of RN using the adaptive threshold. Abrupt changes in T and BC indicate transitions between different climate regimes. Dry events in Lake Lisan (cooling of the higher latitudes) are marked by blue bars and two interstadial peaks (warming) by orange bars. The gray shaded band is the 90% confidence interval for the networks measures.

4.4 Summary

I have introduced a novel method to chose the recurrence threshold adaptively and compared it with the constant threshold selection technique. The selection of recurrence thresholds for recurrence plots and recurrence networks is a crucial step for these techniques. So far, the threshold had to be chosen arbitrarily, taking into account different criteria and application cases as well as requiring some expertise. Here I have proposed a novel technique to determine such a threshold value automatically depending on the time series. Such an adaptive threshold is directly derived from the topology of the recurrence network. It is selected in such a way that the recurrence network does not have unconnected components. I have discussed transitivity and betweenness centrality measures of the complex network approach. Both measures are related to the regularity of the dynamics.

I have compared the novel adaptive threshold selection with the arbitrarily selected threshold by applying them to the logistic map. Although both methods distinguish the dynamical regimes clearly, the adaptively chosen threshold approach detects much more bifurcations, in particular such as period doubling. Such bifurcations are important characteristics of nonlinear dynamics, since these bifurcations route to chaos via period doubling.

Moreover, I have used our approach to investigate a palaeoclimate proxy record from the palaeolake Lisan representing the climate variability in the near East between 27 and 18 cal ka BP. Both transitivity and betweenness centrality measures clearly identified transitions between wet and dry (and vice versa) periods by an abrupt decrease of dynamical regularity, perhaps due to a reduced solar influence. The new method identified some transitions which have not been know so far from the literature and require further investigation, e.g., by analyzing other proxy records from this region. By choosing the adaptive threshold, I have been able to identify the transitions more clearly than by using the arbitrary selected threshold approach.

Chapter 5

The Transformation-Cost Time Series

Time series analysis is an important subject and there are many benefits of it such as understanding the past, estimating the future, evaluating or comparing nonlinear dynamics. According to these different goals, many techniques have been developed over the years. Generally, these methods are introduced for regular time series. However, there are some challenges in real-world time series, e.g, cumulative trends or irregular sampling. In order to deal with these difficulties, many preprocessing methods are introduced such as filtering, detrending or interpolation. Some of these techniques transform an irregular time series to a regular one (ready to processing) while the time series is deformed and loses a lot information.

In this chapter, I discuss a novel approach called 'the TrAnsformation-Cost Time-Series' (TACTS) that can be used to analyse irregularly sampled time series with less damages to the data. This approach is developed together by me and my colleague Ibrahim Ozken. We contributed equally to the research presented. The TATCS method was published in Physical Review E (Ozken et al., 2015). I demonstrate the potential of TACTS on prototypical examples of discrete and continuous time series as well as on a paleo-climate time series. In the last part of the chapter, I present very interesting findings on the relationship of the Holocene East Asian – Australian summer monsoons with TACTS and recurrence plots. This last application, Sec. 5.4.2, follows my recently submitted work (Eroglu et al., 2015).

5.1 Challenges of Real-World Data Sets

5.1.1 Cumulative or Functional Trends in Data Sets

In real-world applications, trends on data are a common problem. For instance, if one applies recurrence-based approaches on a time series with such trend, the data needs to be detrended, otherwise the system will not be recurrent. There are different approaches to detrend a data such as removing linear trends from the data or separating the data into small windows, and subtract the window's mean from each point of associated segments (Peng et al., 1994; Ossadnik et al., 1994). A time series is an observation of a system, in other words, it represents a feature of a system. For instance, let x be a position variable in a one-dimensional motion, and v the velocity of the system. If the time sampling is known, it is easy to transform between these two variables, i.e., these two parameters are dependent on each other over time.

An observation of the time series of only one variable, x or v, is sufficient to describe the system. Let me further exemplify this on a simple signal by a superposition of three sinusoids with a trend parameter a,

$$x(t) = at + \sum_{i}^{3} \sin(\omega_i t + \rho), \tag{5.1}$$

where $\omega_i = \frac{2\pi}{T_i}$ is the frequency of the ith sinusoid with $T_i = [18, 21, 41]$ and ρ is a random number from a uniform distribution, $\rho \in (0, 2\pi)$. Eq. (5.1) gives us a simple trended stochastic process such as in paleoclimate time series. Now I can generate another time series depending on the first one by,

$$v(t) = \frac{x(t) - x(t + \Delta t)}{t - (t + \Delta t)} = \frac{x(t) - x(t + \Delta t)}{\Delta t}, \tag{5.2}$$

where v is another time series associated with the same system, and this transformation Eq. (5.2) detrends the data. Now I compare RPs of time series generated by Eq. (5.1) and transformations by Eq. (5.2). Fig. 5.1(a) is reference figure since there is no trend $a = 0$ for both time series (x and v) and their associated RPs are similar.

For increasing the parameter a, the time series of x inclines while the shape of v does not change Fig. 5.1(b,c). Therefore, the RPs of x drift from the corners of the RP to the line of identity, on the contrary the RPs of v stay constant. Now I have shown that the transformation Eq. (5.2) detrends the time series but, of course, another question appears: "Does the analysis of this new time series v really allows us to detect the same transitions as with x?". In order to answer this question, it is possible to apply this transformation on synthetical time series and compare the results from both x and v.

The logistic map is a good candidate to take as an example since its behaviour and transition points are well known from the previous sections. I rewrite the equation of the logistic map again,

$$x_{i+1} = r x_i (1 - x_i). \tag{5.3}$$

Here I use the same parameters and data sets as in the previous chapter (see Sec. 2.2.2). Then it is easy to obtain the associated time series v from Eq. (5.2). In order to compare the analyses of x and v time series, I have used the DET measure of RPs (Eq. (2.20)) since it is a well-suitable measure to find transitions.

In Fig. 5.2 I present the results. The Lyapunov exponent of the logistic map for the selected range of control parameter $r \in [3.5, 4]$ is plotted in Fig. 5.2(a). The determinism is calculated for both time series (x and y) and the results are given in Fig. 5.2(b,c). Comparison with the Lyapunov exponent shows that the determinism of both time series (x and y) tracks the transitions of the different regimes for all regions of control parameter r. All instantaneous drops in the determinism values indicate the transitions between positive–negative regions of the Lyapunov exponent. The transitions are highlighted with dashed lines in Fig. 5.2. Therefore, the analysis

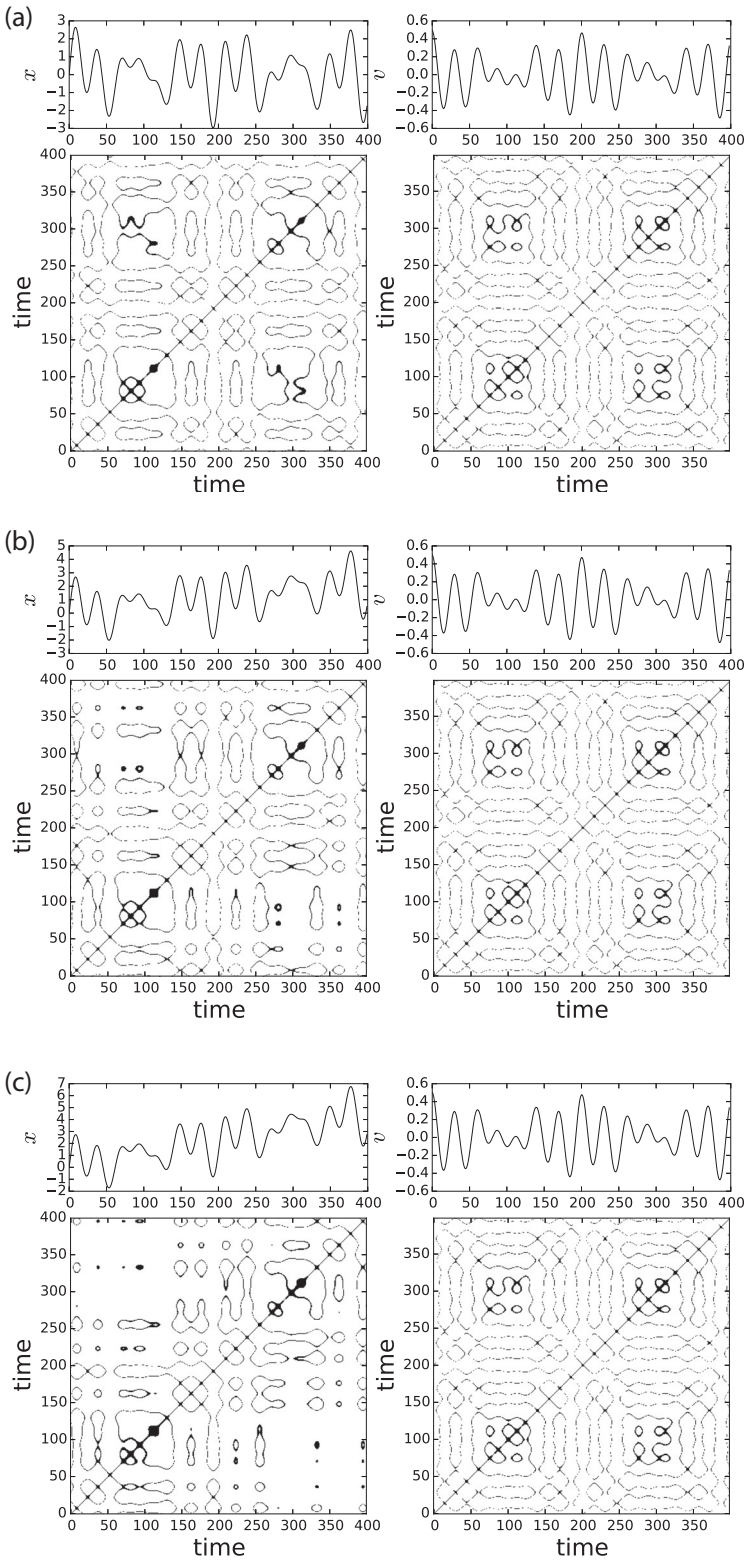

Figure 5.1: Generated time series x with Eq. (5.1), transformations from x to v with Eq. (5.2) for different trend parameters a and their associated RPs , (a) a= 0.0, (b) a = 0.003, (c) a = 0.006.

shows us that using the time series v instead of x maintains the underlying dynamics of the system as well.

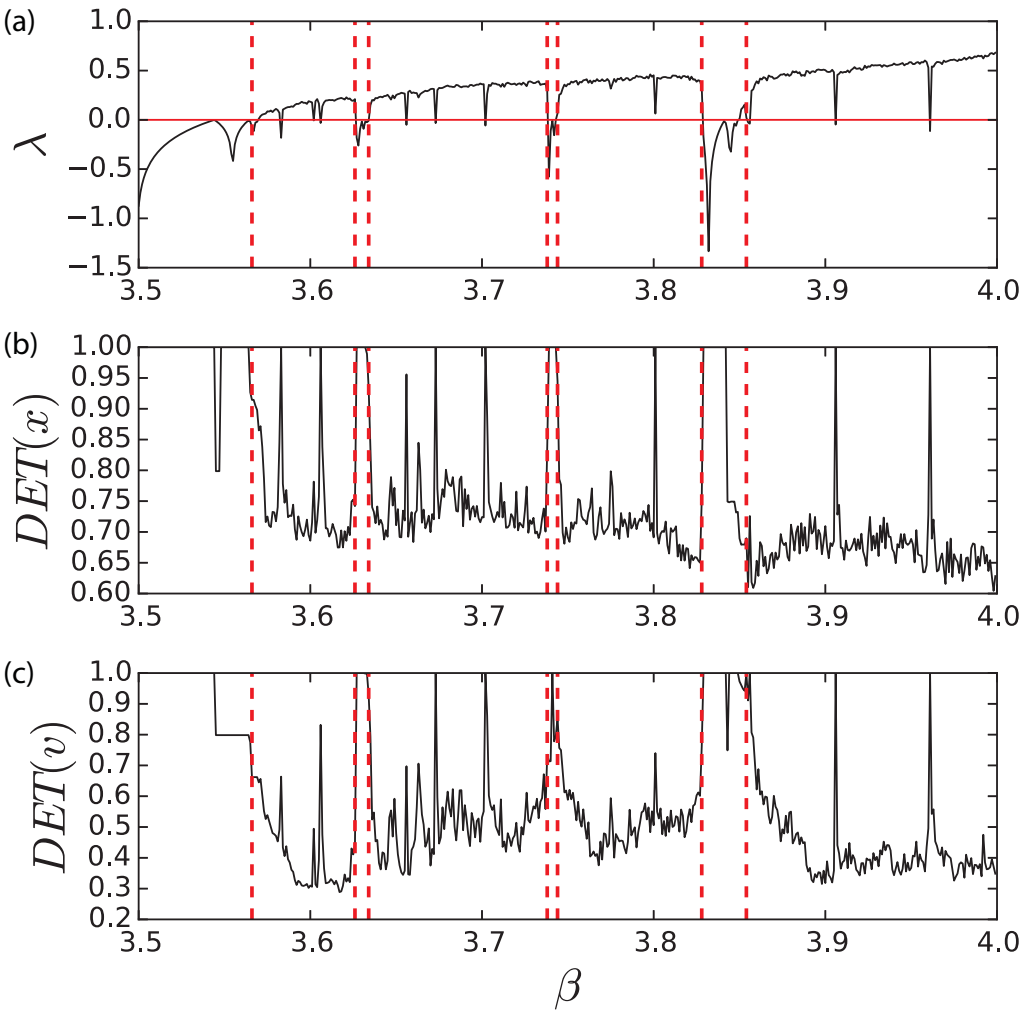

Figure 5.2: Comparison between (a) the Lyapunov exponent and the determinism values for (b) the logistic map (x) and (c) the transformation (v) data. Information about the transformation in detail can be found in the text. Dashed lines indicate dynamical regime transitions and the determinism shows critical points clearly.

The Rössler system is the second application and it is a good example to test the transformation on flows (continuous systems). The Rössler attractor is studied in a previous chapter, I will use the same parameters for the consistency Sec.3.3.2 ($b \in [0, 2]$, $a = 0.2$, $c = 5.7$). The system is given by

$$\dot{\boldsymbol{x}} \quad = \quad \boldsymbol{f}(\boldsymbol{x}) \tag{5.4}$$

$$\left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) = (-y - z, x + ay, b + zc). \tag{5.5}$$

For each b value, I create a time series of the length $N = 6 \times 10^5$ and discard the transient responses by using only the last 5×10^3 data points for the following analysis. The vector transformation is applied analogously to Eq. (5.2). Then $\boldsymbol{x}$ and $\boldsymbol{v}$ are analyzed using the determinism. In Fig. 5.3, the transition points are shown by dashed lines. The determinism of both data sets shows transitions as expected and it is easy to obtain from these results that the transformed time series carry underlying information of the system. In order to analyze a system, it is possible to use both time series.

Therefore, in order to detrend a time series, one can use this transformation without loosing any information from the system. Now I will use this approach on a different challenge which is *irregularly sampled time series*.

5.1.2 Irregularly Sampled Data Sets

There are numerous methods that can be used to detect dynamical regime changes in regularly sampled times series, i.e. the time resolution $\Delta t = t(s_{i+1}) - t(s_i) = \text{const}$ $\forall i \in [0, N-1]$, e.g. (Livina and Lenton, 2007; Trulla et al., 1996; Malik et al., 2014). However in several disciplines like astrophysics and earth sciences a constant sampling cannot be ensured. For instance, a paleo-climate data set from the Dongge Cave, China is given in Fig. 5.4(a) with the histogram of time differences (b). Obviously, the sampling time is not constant along the data. Therefore, regular interpolation as a preprocessing step is often applied, but this might lead to a bias of the results (Rehfeld et al., 2011; Rehfeld and Kurths, 2014). For example, interpolation leads to a positive bias in autocorrelation estimation (and, thus, an overestimation of the persistence time) and a negative bias in cross correlation analysis (Rehfeld et al., 2011).

Here I explore a different approach that can be used for irregularly sampled data without interpolation. Focussing on the palaeo-climate time series, I am in particular interested in regime changes. Such palaeo-climate proxy records show a very erratic sampling, with the sampling times being often distributed with high skewness (Rehfeld et al., 2011; Rehfeld and Kurths, 2014). In addition the records are subject to measurement noise, which makes interpolation difficult (Fig. 5.4). Instead of interpolating the time series I determine TACTS between segments of the original time series which results in a new transformation cost time series having regular sampling. This transformation costs time series can then be analysed by established methods. Wanting to detect regime changes I apply recurrence plot analysis being one of the appropriate methods for this purpose (Marwan et al., 2007).

This new approach is based on a measure introduced by Victor and Purpura (Victor and Purpura, 1997) which was further developed in order to transform spike trains to real-valued time series with regular resolution by Hirata and Aihara (Hirata and Aihara, 2009). The idea behind this approach is similar to the fluctuation of similarity (FLUS) method, that is: if the time series is from one dynamical regime, the cost of transformation from one segment to the subsequent one should be similar for each

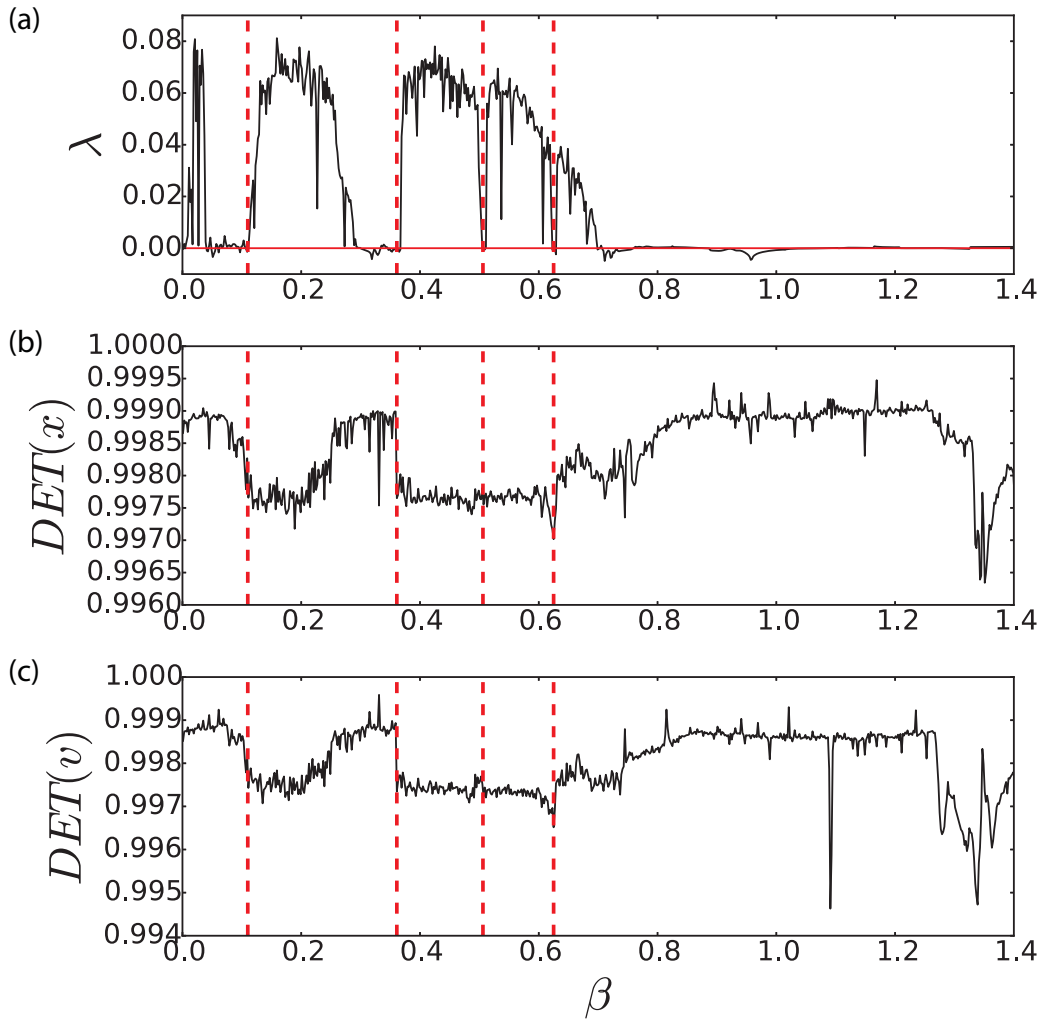

Figure 5.3: Comparison between (a) the Lyapunov exponent and the determinism values for (b) the Rössler attractor (x) and (c) the transformation (v) time series. Dashed lines indicate dynamical regime transitions and the determinism shows critical points clearly.

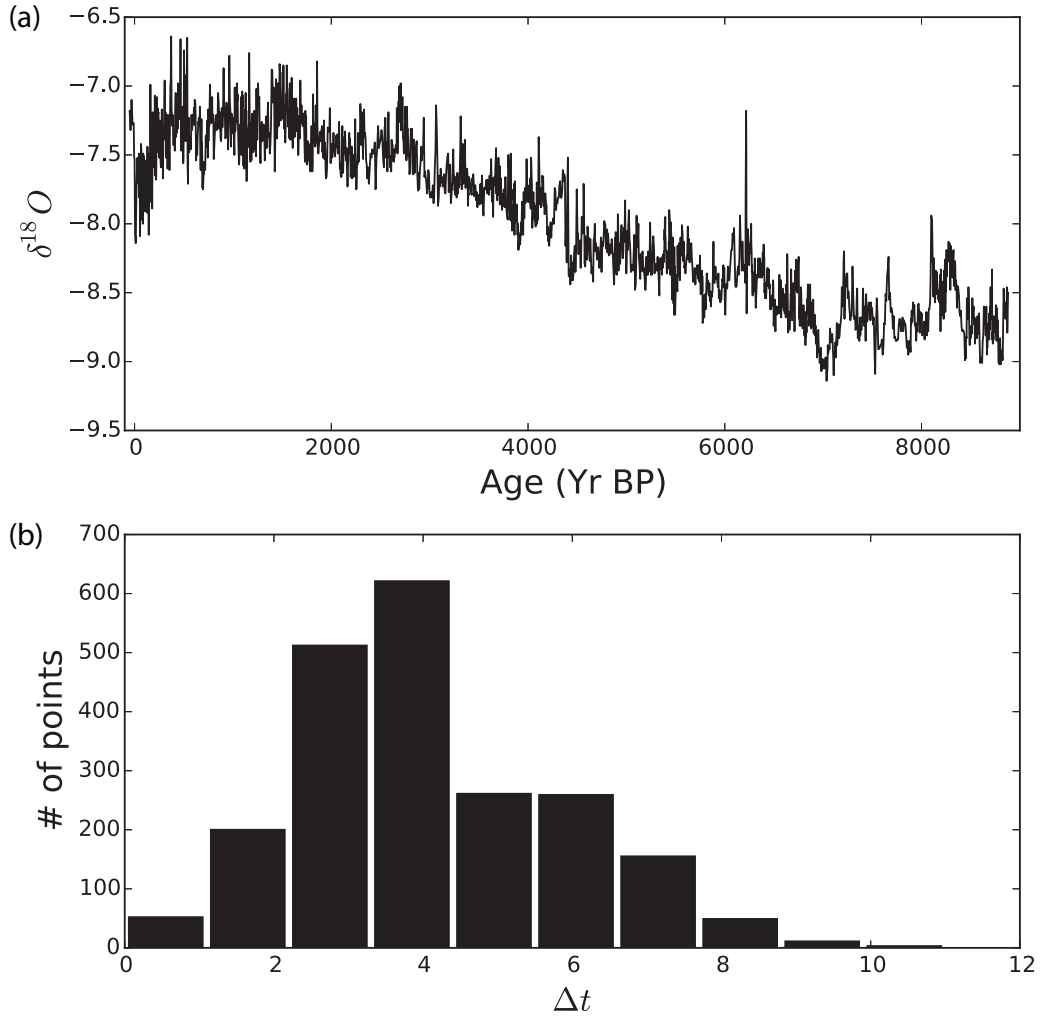

Figure 5.4: (a) $\delta^{18}O$ palaeo-climate proxy from Donnge Cave, China. (b) The histogram of the sampling time. In general, palaeo-climate proxies are not regularly sampled and traditional time series analysis techniques such as power spectrum, recurrence based methods etc. are not directly applicable.

segment of the data (Malik et al., 2014). Since, for example in palaeo-climate proxy records, one may not have any knowledge about the current control parameters, we compute the cost of transforming one segment into the following one. Dramatic changes in the cost time series indicate a change in the underlying dynamics.

While recurrence plot based quantification is not directly applicable to irregularly sampled time series, we show that identifying regime changes in the dynamics of the system becomes possible by combining the TACTS approach with recurrence analysis.

5.2 Metric Distances

5.2.1 Spike trains

In time series analyses, an important issue is to calculate the distance between two data patterns. This distance problem frequently occurs for example in recurrence plots (RP) (Marwan et al., 2007), the estimation of the maximum Lyapunov exponent (Michael T. Rosenstein and Luca, 1993), scale-dependent correlations (Rodó and Rodriguez-Arias, 2006), data classification (Sakoe and Chiba, 1978) or correlation dimension estimations (Grassberger, 1983b). When doing these kinds of analyses with the data having an equidistant time resolution, the Euclidean distance is typically used. However, in applications that generate data with non-regular sampling such an approach is not directly applicable. Such data sets include almost all palaeo-climate observations which can have a very erratic –almost random– time resolution. One way to deal with such data sets is to interpolate them. Such an interpolation will not only fill the gaps but replace real measurements with new interpolated data points closeby that have regular sampling. But this is often not the optimum method, since the subsequent analysis will be typically biased (Rehfeld et al., 2012). Moreover these interpolated values have a higher uncertainty than the measured data points they replace.

For the dynamics of firing neurons, Victor and Purpura (Victor and Purpura, 1997) showed that the spike time distance is a useful method that applies to irregularly sampled data sets. The basic idea of this method is a distance metric that provides information of how easily one data segment can be transformed into another one. To transform one segment of data into another segment, three elementary operations are required: adding or deleting a data point and moving the data point to a different time. Using associated costs for these elementary operations, an optimal data transformation will be achieved if the cost of the transformation is minimised. I will illustrate this method for spike train data below before introducing our modified method for continuous data that determines TACTS.

Consider the metric D as a mapping of two pairs of spike trains or data segments, say S_a and S_b, onto a real value. In order for D to be metric, it must satisfy the following three conditions since D is a generalised distance:

- $D(S_a, S_b) \geq 0$ (positive)

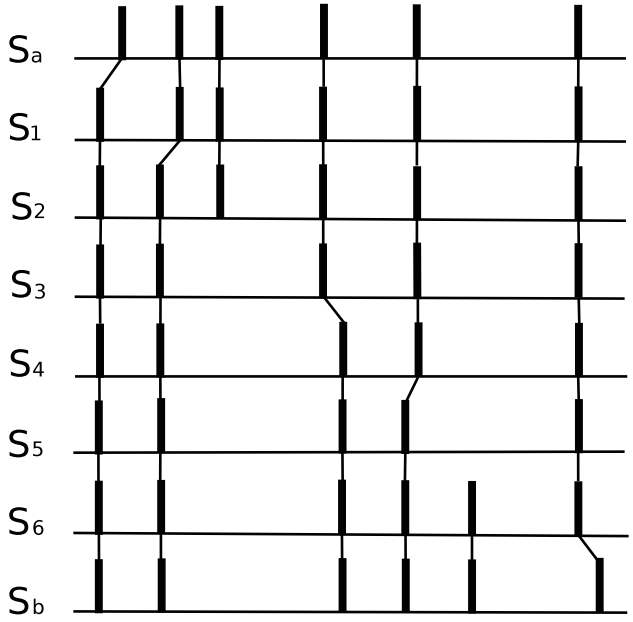

Figure 5.5: Illustration of the transformation of S_a to S_b. In total S_a undergoes seven steps numbered $S_1, S_2, \ldots, S_6$. Note that $S_7 = S_b$. The path shown is a minimal-cost path and all the steps are elementary steps, like moving a spike or deleting and creating.

- $D(S_a, S_b) = D(S_b, S_a)$ (symmetric)
- $D(S_a, S_c) \leq D(S_a, S_b) + D(S_b, S_c)$ (triangle inequality)(Victor and Purpura, 1997)

In Fig. 5.5 I give one illustrative example how the elementary operations transform the spike train S_a into S_b (Victor and Purpura, 1997). When I transform S_a to S_b, the total cost is the sum of the elementary step costs. As one can see the transformation of S_a to S_b requires 7 distinct steps. Step 1, 2, 4, 5 and 7 move one spike to a different time point, while step 3 deletes one spike and in the 6th step one spike is created. Assigning costs to each of these operations and pairwise checking the segments, Victor and Purpura (Victor and Purpura, 1997) analysed different types of spike time series. They chose the cost of deleting and adding to be the same $p_d = p_a = 1$, while moving a spike is proportional to the time distance that the data point is moved from t_a to t_b: $\lambda_0 |t_a - t_b|$. Clearly the parameter λ_0 is a frequency with the unit Hz.

5.2.2 Marked data

So far I only considered spike trains as they are apparent in the analysis of brain dynamics data. Suzuki *et al.* (Suzuki et al., 2010) have extended this method for continuous marked data and I follow their approach. The continuous data set is

transformed to an event time series that is very similar to spike trains, and allows events to have different amplitudes. The transformation from continuous data to an event time series is done by the system itself that is event occurrence as earthquakes, data quality as palaeo-climate time series, extreme events as crisis in finance etc.

To calculate the transformation cost time series I determine the cost for the transformation of one segment into the other for two successive segments of a time series. For a single transformation, this cost is a generalised distance between these two segments. Therefore, as a distance, the cost must be a positive number, symmetrical (i.e. transforming the first into the second is the same as transforming the second into the first), and must satisfy the triangle inequality.

In Fig. 5.6 I give an illustration of how to perform this transformation. Recall that the transformation is done by 3 elementary steps: (i) shifting an event in time; (ii) changing the amplitude of the event; and (iii) creating or deleting an event. The figure outlines the steps required to transform the top time series segment into the bottom one. This transformation consists of 7 elemental steps. Moves 1 and 2 move the first and second event to the right and, in addition, adjust their magnitude, i.e. a combination of the two elementary steps (i) and (ii). In move 3 the last event is deleted (that is, elementary step (iii)). As one can see it takes 4 additional elementary steps (combinations of (i) and (ii)) to transform the starting time series into the target time series.

The cost associated with each transformation ($S_a \to S_b$) is given by:

$$\begin{aligned} p(c) = \quad & \textstyle\sum_{(\alpha,\beta)\in C} && \{\lambda_0 |t_a(\alpha) - t_b(\beta)| + \\ & \tfrac{1}{m}\textstyle\sum_{k=1}^{m} && \lambda_k |L_{a,k}(\alpha) - L_{b,k}(\beta)|\} + \\ & \lambda_S(|I| && +|J| - 2|C|), \end{aligned} \tag{5.6}$$

where I and J are a set of indices of the events in starting set S_a and the final set S_b, respectively. These sets – S_a and S_b – consist of the events in the two time series segments. The first summation quantifies the cost associated with shifting events in time. I sum over the pairs $(\alpha, \beta) \in C$, where the set C comprises the points that need to be shifted in time. α and β are the αth event in S_a and βth event in S_b. The coefficient λ_0 is the cost factor for time shifts. The second summation calculates the cost due to changing the amplitude of events. This involves the difference $|L_{a,k}(\alpha) - L_{b,k}(\beta)|$, where $L_{a,k}(\alpha)$ is the amplitude of the αth event in S_a. The parameter λ_k has the unit of amplitude^{-1} and the sum is over the different components of the amplitude. That is, if one is dealing with one dimensional data $m = 1$, while for a three dimensional phase space m would be three. The last terms in the cost function deal with the events not in C which have to be added or deleted. Note that $|\cdot|$ denotes the size of the set and λ_S is the cost parameter for this operation. Suzuki *et al.* omitted this parameter, since they chose a unit cost for such an operation (Suzuki et al., 2010).

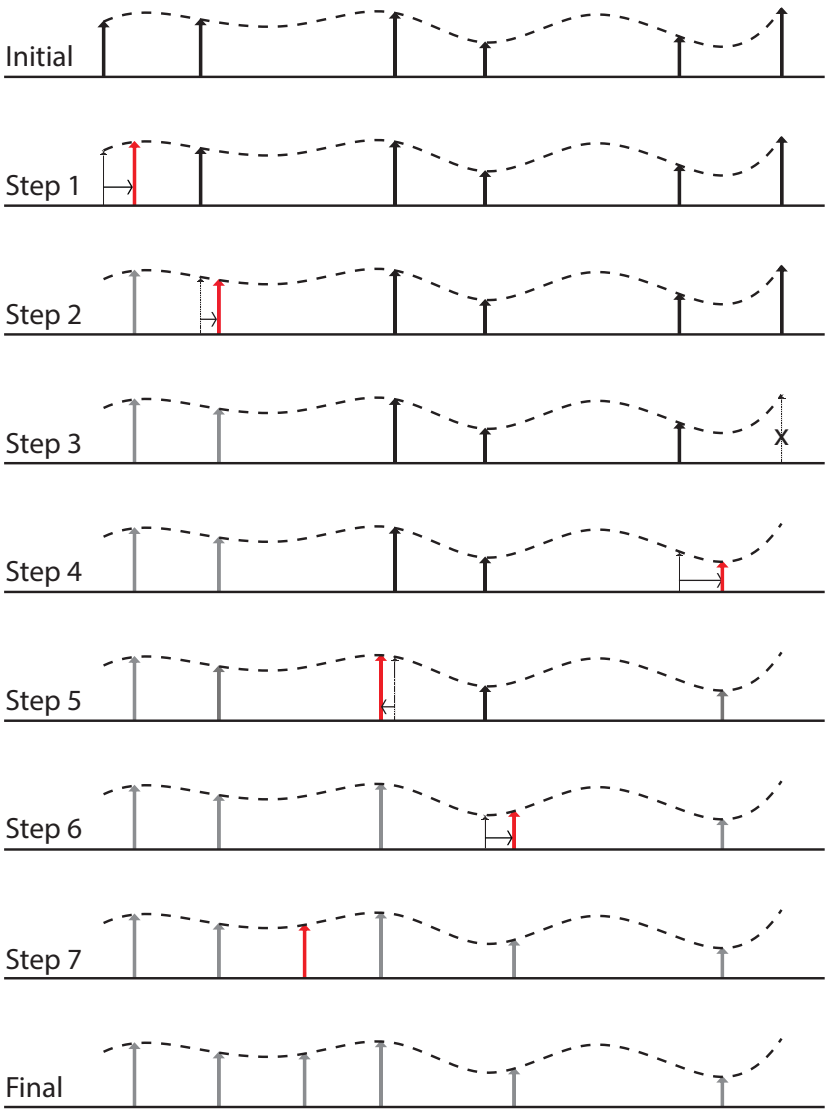

Figure 5.6: Illustration of the transformation cost time series method. The initial time series segment (top) is transformed into the final time series segment (bottom) in seven steps. Note that after seven steps the segment is identical to the final target time series. The steps 1, 2, 4, 5, 6 are combinations of the elementary operations (i) time shift and (ii) adjusting the amplitude (first two terms in Eq. (5.6)) while in step 3 one event is deleted and therefore the (iii) elementary operation was applied (last term in Eq. (5.6)).

I determine the cost factors λ_0, λ_k based on the time series at hand:

$$\lambda_0 = \frac{M}{\text{total time}} \tag{5.7a}$$

$$\lambda_k = \frac{M-1}{\sum_i^{M-1} |L_a - L_b|}, \tag{5.7b}$$

where M is the total number of events in the time series. Note that λ_0 is the mean event frequency and λ_k is the inverse of the average amplitude difference.

The cost factor λ_S is an optimisation parameter. I constrain $\lambda_S \in [0, 4]$ and explore the costs of deleting or adding an event to our time series. If our time series consists of $n+1$ segments of equal length, I can calculate n costs for each individual transformation of the segments. Assuming that the costs are statistically independent, the central limit theorem indicates that the distribution of the costs should be a normal distribution. In particular, when dealing with non-stationary data we find that changing λ_S such that the distribution becomes normal greatly improves the skill of our time series analysis method.

Assume a time series: $X = (x_{t_1}, x_{t_2}, x_{t_3}, ..., x_{t_N})$, where N is the number of points and $t_2 - t_1 \neq t_3 - t_2 \neq ... \neq t_N - t_{N-1}$. I divide the time series to a set of segments W which have equal size. After this, we have n equal windows $W_1, W_2, ..., W_n$ and I determine the transformation costs $p(W_1, W_2), p(W_2, W_3), ..., p(W_{n-1}, W_n)$ for all sequence windows. This leads to a new equidistant time series which is our transformation cost time series. By using a RP analysis we detect regime changes in the underlying dynamics.

An intuitive understanding of the transformation cost time series method is based on an interpretation of the cost function (5.6) and the cost coefficients $\lambda_{0,k,S}$ (5.7). As mentioned above $\lambda_{0,k}$ are the average amplitude and the average event frequency while λ_S penalizes changes in time and amplitude of an event that are large. These coefficients weight the local difference between the event pairs in our cost function (5.6). Therefore I can perceive the cost function as balancing the time and amplitude differences of the events in the two segments (S_a and S_b in Fig. 5.5) versus deleting and recreating all events. If I analyze several segments resulting from a regular dynamics, the local difference between the segments will be bounded and the cost time series will show some regularity. If the underlying dynamics on the other hand is erratic, the local difference between the segments can be large and consequently the cost function shows no obvious regularity (this property is also used for the FLUS method (Malik et al., 2014)). Differencing a (regular) time series (applied, e.g., as a high-pass filter), $x_t - x_{t-1}$, is a special form of this approach. To measure regularity in our transformation cost time series I apply recurrence quantification analysis.

5.3 Applications to Prototypical Models

In real world applications, especially in the palaeo-climate data sets, time series are not equidistantly sampled. In order to deal with this kind of difficulties, I have

created prototypical irregularly sampled models from a discrete (the logistic map) and a continuous (the Rössler attractor) systems. Moreover, for the high likelihood of noise in real world applications, I have added noise into the study on the logistic map.

5.3.1 Logistic map

As the first application I analyze data from the logistic map, which is defined as:

$$x_{i+1} = r x_i (1 - x_i), \tag{5.8}$$

for $r \in [0, 4]$.

It has been shown (Trulla et al., 1996; Marwan et al., 2009; Eroglu et al., 2014a; Eroglu et al., 2014b; Marwan et al., 2002) that analyzing RPs is an efficient method to detect the regime transitions in the logistic map's dynamics. I will analyse the dynamics and its transitions in a control parameter range of $r \in [3.5, 4]$. For this investigation I sample the control parameter range with a step size of 0.001 and calculate a time series of 3000 iterations for each control parameter value. The first 1000 points are deleted to discard transients, resulting in time series consisting of 2000 points that have been used for all analysis of the logistic map in this application.

I investigate the performance of the new method for non-equidistant sampled data by deleting randomly 100 ($\gamma = 5\%$), 200 ($\gamma = 10\%$), 300 ($\gamma = 15\%$) or 400 points ($\gamma = 20\%$) from the original time series. For all time series I choose a segment size of four time steps. This size can still capture changes in the underlying dynamics even for $\gamma = 20\%$ but also results in a long enough transformation cost time series that can be analysed using RP. I determine the transformation cost for each window pair in the data set using $\lambda_{0,k}$ given from Eqs. (5.7) and optimised λ_S as outlined in Sec.5.2. The value of λ_S depends on the particular γ and does decrease with increasing γ. While one could argue that for each chaotic regime one should use a different λ_S, I chose to determine only one value from the time series generated at $r = 4$. There are two reasons for this particular choice: (i) using one λ_S for all time series resembles the situation when no additional information about the control parameter is available and (ii) it shows that our method does not crucially depend on the choice but is stable even if λ_S is close enough to the optimum value. For our different γ levels I determined the optimal λ_S to be: $\lambda_S = 1.07$ ($\gamma = 5\%$), 1.04 (10%), 0.95 (15%) and 0.93 (20%).

This optimisation results in $\lambda_S \approx 1.00$ on average and therefore is similar to the original values used in (Victor and Purpura, 1997) and (Suzuki et al., 2010). This transformation cost time series is then used in the RP to calculate the determinism Eq. (2.20) with $\varepsilon = 0.08$ for all r values considered. The ε value needs to be sufficiently small (Marwan et al., 2007) and $\varepsilon = 0.08$ is adequately suitable for our phase space of the transformation cost.

In Fig. 5.8 I present the results. Panel (a) shows the Lyapunov exponent calculated from the time series. In panel (b), the determinism calculated for the time series is

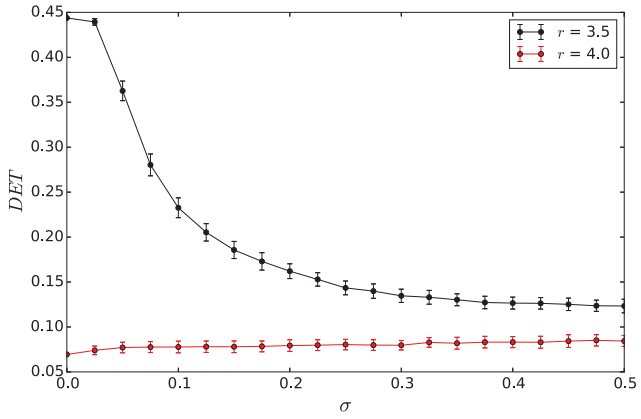

Figure 5.7: Determinism against noise level σ: $\gamma = 10\%$ and $r = 3.5$ (periodic, black line) and $r = 4$ (chaotic, red line) .

shown for increasingly irregular sampling. A comparison with the Lyapunov exponent shows that the determinism tracks the transitions of the dynamics for all data sets. All abrupt drops from positive Lyapunov exponent to negative ones are apparent in the determinism measure. These drops are indicated with dotted lines in Fig. 5.8. One clearly sees that randomly deleting points leads to a distinct drop in the determinism. While this drop is most pronounced when comparing the reference line ($\gamma = 0\%$) with $\gamma = 5\%$ for higher γ values the determinism does not decrease as much. This is due to the fact that any deleting will disconnect the long diagonal lines that contribute mainly to Eq.(2.20). Naturally the results are most conclusive for $\gamma = 0\%$ but even for $\gamma = 20\%$ it is possible to identify the bifurcations in the logistic map and successfully detect the changes in the dynamics.

In Fig. 5.8 panel (c), showing that this method cannot only detect changes in the dynamics for irregular sampling but is also stable if the data is additionally compromised by measurement noise. I added Gaussian white noise ($\langle\xi\rangle = 0$ and $\langle\xi(t)\xi(t')\rangle = \delta(t - t')$) and the results shown are for a noise level of $\sigma = 0, 0.05$ and 0.1 ($\gamma = 5\%$ and 10%). Note that the σ is scaled relative to the variance of the time series. Again one can clearly see that for all noise levels considered, our method is able to identify the changes in the dynamics and closely follows the Lyapunov exponent.

To further investigate the stability of the method I investigate two time series of 2000 points corresponding to $r = 3.5$ (periodic dynamics) and $r = 4$ (chaotic dynamics) using $\gamma = 10\%$ and increase the noise standard deviation σ in steps of 0.01. Fig. 5.7 shows that one can clearly distinguish between periodic dynamics and chaotic dynamics even for high noise levels. The bullets give the average determinism for a 100 time series ensemble, while the error bars show the standard derivation of the ensemble. It should be noted that for these two extreme chases– periodic and chaotic – the bands are clearly separated for the whole σ range considered but to be on the safe side I would not recommend analysing data with more than $\sigma = 0.2$.

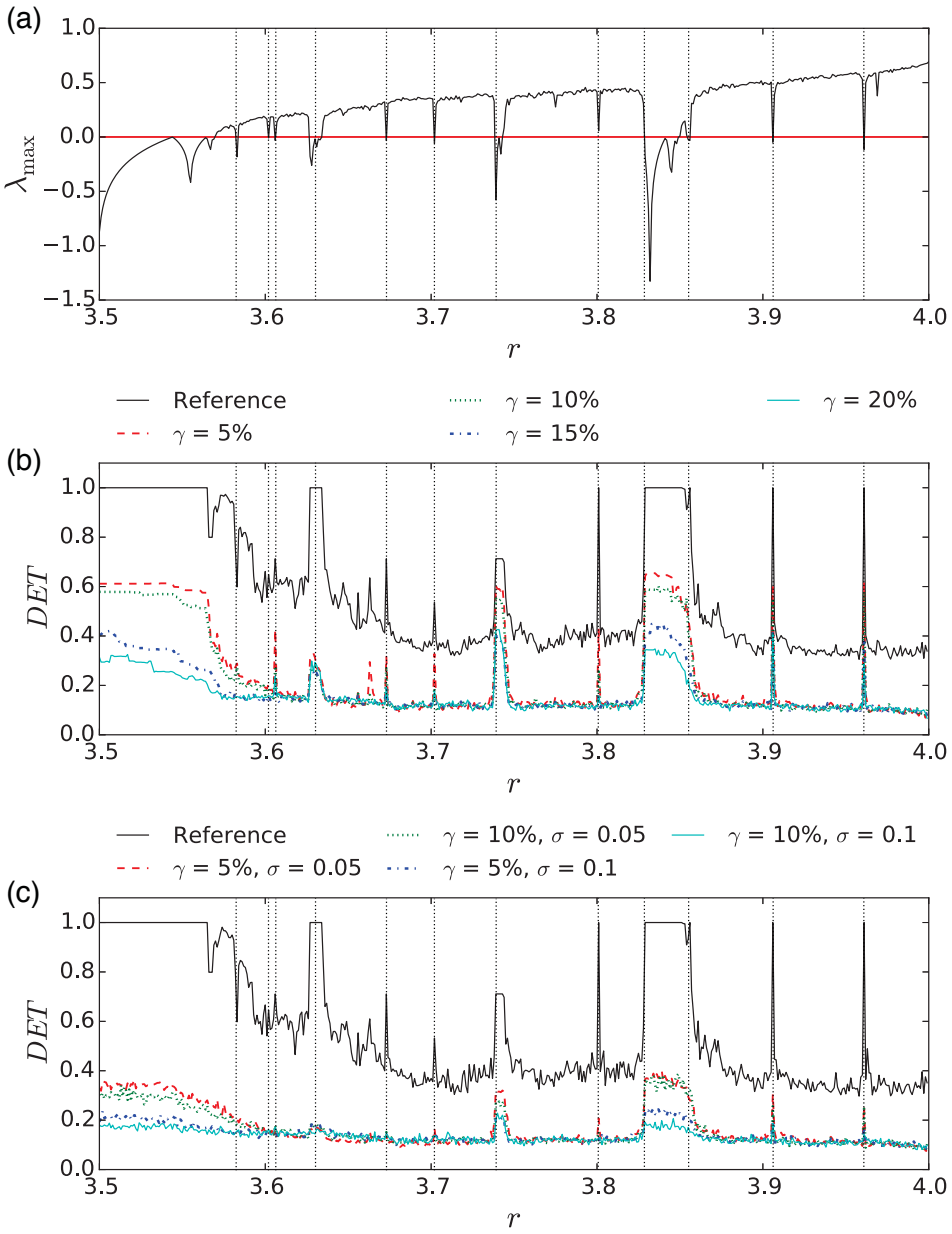

Figure 5.8: RQA analyses for Logistic Map: (a) Lyapunov exponent of the Logistic Map; (b) - (c) Determinism calculated from TACTS for (b) various levels of deleting; and (c) two measurement noise levels ($\sigma = 0.05$ and 0.1) and for two different rates of irregularity ($\gamma = 5\%$ and 10%). The reference is the analysis from the original time series without any deformation.

Nevertheless this method is quite stable even if corrupted by measurement noise.

5.3.2 Rössler Attractor

In order to mimic irregular sampling, I consider again the continuos Rössler system:

$$\left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) = (-y - z, x + ay, b + z(x - c)), \tag{5.9}$$

where a, b and c are parameters. In this application I chose $a = 0.2$ and $c = 5.7$ and vary $b \in [0, 1.4]$ with a resolution of $\Delta b = 0.01$. To achieve an irregular sampling, I use the maximum map $\tilde{Y}$ of the y-component, which offers a natural way to get non-equidistant sampled event time series in the chaotic regime as well as in the windows of higher periodicity. For this investigation I generate a long time series via using the 4th order Runge-Kutta method with $\Delta t = 0.01$ sampling rate. Then I neglect any transient behavior and consider 5000 maxima for each control parameter value b. Using a window size of 3500 time units I calculate TACTS with parameters $\lambda_{0,k}$ determined by Eqs. (5.7) for all b values and then optimize λ_S such that the cost distribution is normal.

In addition to the irregular sampling real world data sets may contain measurement uncertainties. While it is known that this method performs well even with measurement noise added to the dynamics (c.f. Fig. 5.7), the Rössler system offers an additional opportunity to test for a different kind of uncertainty. As mentioned above palaeo-climate proxy records are often Gamma distributed (skewed distribution) in the time domain (Rehfeld et al., 2011; Rehfeld and Kurths, 2014). To let the data reflect this, I first create a cubic interpolated maximum time series resulting from $\tilde{Y}$ acting on y. Then I choose Gamma distributed time events at which I sample the interpolated time series to create a new time series with higher uncertainty. The two step process is illustrated in Fig. 5.9 for two chosen skewness values. For the analysis I use the skewness of the Gamma distribution as 0.3, 0.5, 1.0 and 2.0. Therefore I generate four additional time series that I analyze by determining TACTS and determine our RQA measure DET with $\varepsilon = 0.05$.

The results for the five time series are shown in Fig. 5.10. Panel (a) shows the Lyapunov exponent calculated from the continuous sampled y component. In the panel below we see the data for the five different time series I considered. Clearly the technique is able to identify the dynamical regime changes (dotted lines in Fig. 5.10) for all data sets. Just like with increasing noise intensity in the logistic map the changes are not so pronounced anymore for higher skewness, but even for a skewness of 2.0, the chaotic regime can be clearly distinguished from the periodic dynamics and it is possible to identify the regime changes associated with the periodic windows.

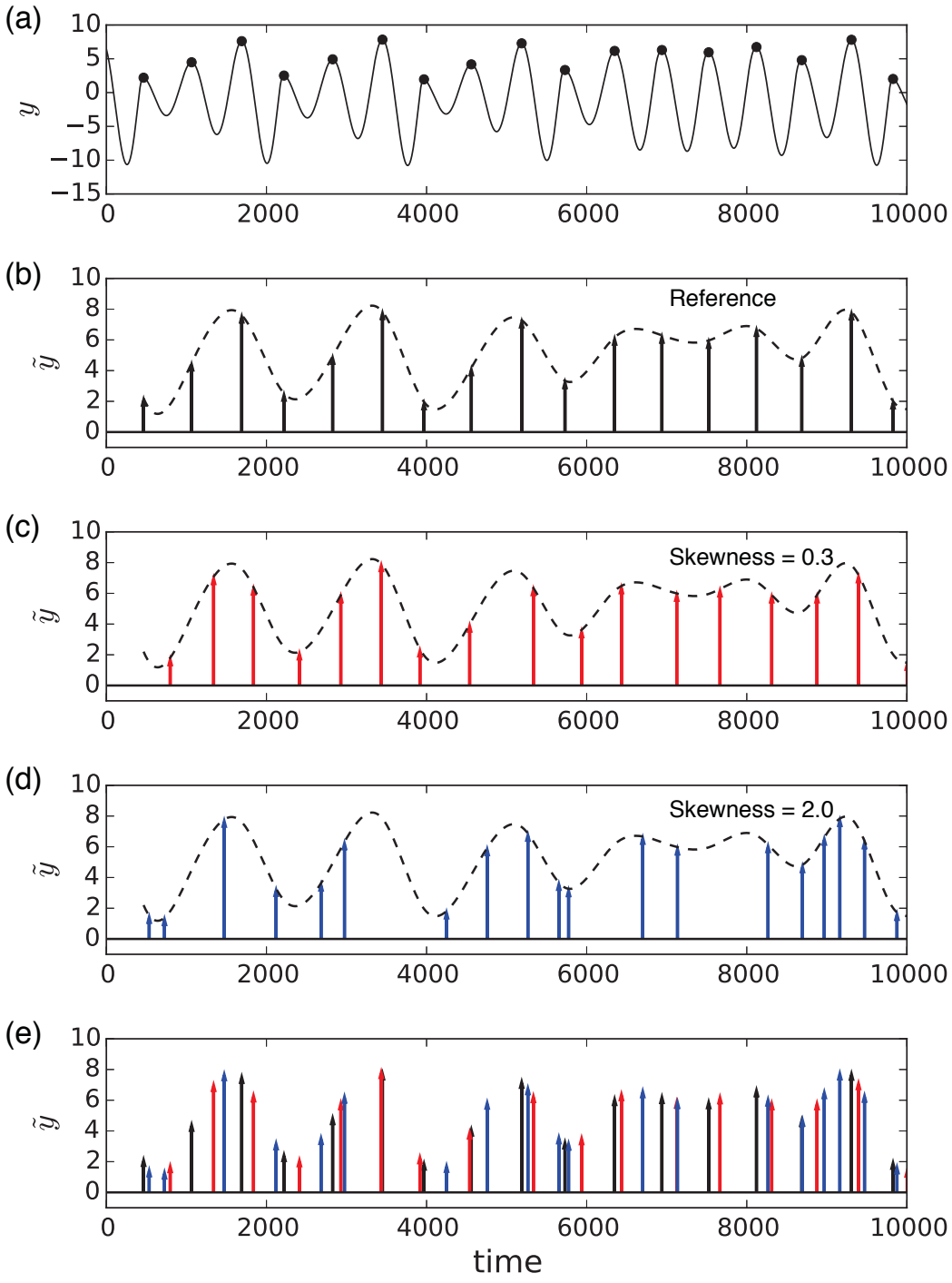

Figure 5.9: The time series of Rössler: Panel (a) shows the y-component with maxima highlighted as bullets; (b) shows the result of the maximum map $\tilde{Y}$ acting on y; (c) shows the time series for a skewness of 0.3 and (d) for a skewness of 2.0. In panel (b) to (d) I show the interpolated time series that I draw from as a black dashed line. Panel (e) offers a comparison between the different time series and highlights their irregularity.

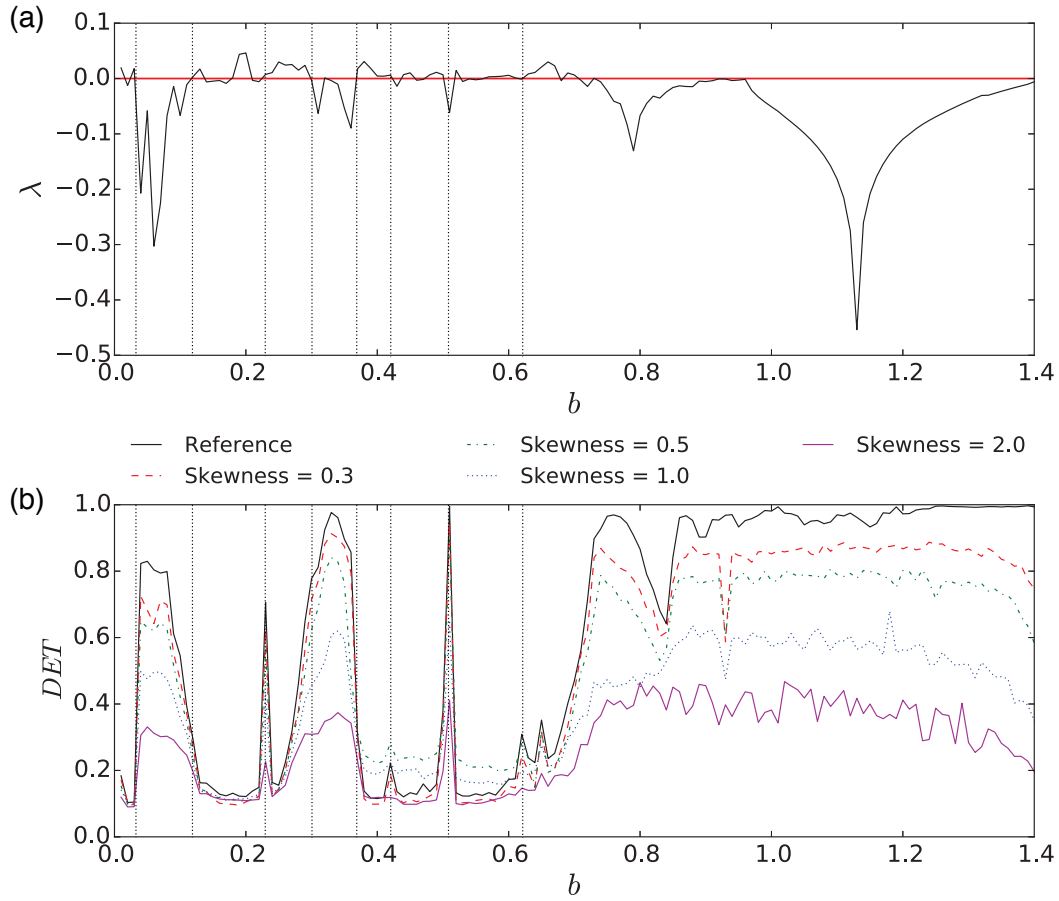

Figure 5.10: RQA analyses for Rössler Map distance time series: In panel (a) the Lyapunov exponent λ is given as a reference over the whole b control parameter range. In panel (b) the determinism DET determined from TACTS is shown for the maximum time series $\hat{Y}$ (black line) as a reference. The other shown data results from the Gamma distributed time series with skewness 0.3 (red dashed line), skewness 0.5 (green dash-dot line), skewness 1.0 (blue dotted line) and skewness 2.0 (pink line).

5.4 Applications to Palaeo-climate Record

So far I have been testing the performance of the transformation cost time series method using prototypical models. While I have been trying to design these numerical examples as realistic as possible – including measurement noise and testing irregular gamma distributed sampling – in real applications the data might have further complications like multiplicative noise and time dependent control parameters. Since I am particularly interested in palaeo-climate applications, I choose speleothem isotope δ^{18}O records as real-world applications.

5.4.1 The Secret Cave, Borneo

First, I choose a speleothem isotope δ^{18}O record from the Secret Cave at Gunung Mulu in Borneo / Indonesia (Fig. 5.11) (Carolin et al., 2013) . This particular record is a proxy for the Indonesian-Australian monsoon, since δ^{18}O is an indicator for precipitation. I analyzed the last 62,000 years of this proxy record. Note that the full record is around 100,000 years long, but beyond 62,000 years the data is too sparse and contains too many gaps to give any useful information.

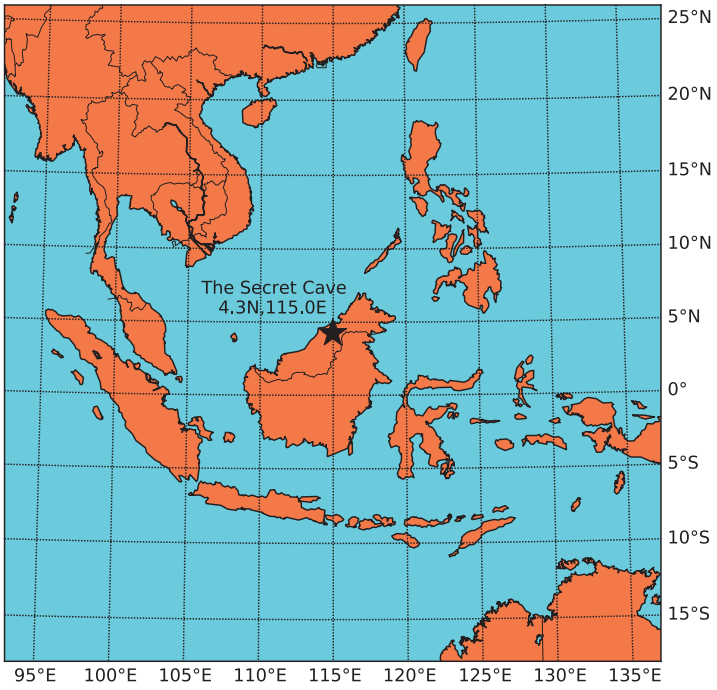

Figure 5.11: Map indicating the location of The Secret Cave, Borneo, Indonesia

In the part of the record analyzed there are ∼1200 data points. The time between two measurements follow a Gamma distribution and the skewness is 4.9. I use a segment size of ≈ 210 years to calculate our transformation cost time series. The parameters $\lambda_{0,k}$ are determined by Eqs.(5.7) and I optimize $\lambda_S = 1.07$. To detect dynamical transitions in TACTS, I need to apply a slightly different form of the RQA (Marwan et al., 2013), since in the palaeo-climate data I expect a temporal variation of the control parameter. I therefore adopt a sliding window method and consider thirty data points of TACTS as our window size. Note that 30 data points in TACTS correspond to approximately 100 to 140 points in the original proxy record and cover about 6200 years in real time. The length of a window (∼ 6200 years) is a suitable interval to detect the regime transitions in palaeo-climatology. DET is determined for each window of the time series one by one as the window slides over the time series with 90% overlap. I used $\varepsilon = 20\%$ of the standard derivation of the data in the particular window. Not only does this method allow us to deal with data that shows regime changes due to control parameter variations, it also gives us a way to determine the statistical significance of DET via the method of bootstrapping as outlined in (Marwan et al., 2013). The basic idea is that the dynamics of the system does not change over time. The bootstrapping test will allow us to judge whether the found variability of the measure DET is significantly different from an unchanged dynamics, i.e., whether a regime transition occurs.

In Fig. 5.12 I present the results of the analysis together with the original proxy record. Outside the light red band DET can be considered to indicate a dynamics different from the 'normal behaviour' with 90% confidence. As one can see, the RQA-determinism indicates several distinct regime changes in the time series. Quite pronounced are the regime changes that coincides with the known Heinrich events (H1 to H6). During these Heinrich events large quantities of fresh water was introduced into the Atlantic via melting ice-bergs (Bond et al., 1992) and it is apparent, that these events impacted also on the Monsoon dynamics over the Maritime Continent (Pausata et al., 2011), leading to very low DET. Similarly the Younger Dryas, a period of cool climate in the Northern Hemisphere that might have been caused by the collapse of the North American ice sheet (Berger, 1992), is detected by the method. It is noteworthy that in the original work by Carolin *et al.* (2012), H1 to H6 was detected too, but the Younger Dryas coinciding with the Heinrich 0 event (H0), was not detected. Moreover our method allows an objective, quantitative analysis, while Carolin *et al.* rely on the subjective method of matching extreme data occurrences with specific dates.

In the analysis I detect some other significant regime changes (see for example the high DET between H2 and H3 in Fig. 5.12) that have previously not been recognized. Similarly I observe that while all Heinrich events were impacting on the climate significantly, the duration and strength with which they impacted on the monsoon over Indonesia varied according to our analysis. From a palaeo-climatic point of view the monsoon dynamics over the Maritime Continent is quite complex with cold surges from the north impacting on the local precipitation. In addition changes in the landmasses due to rising sea levels generated the Borneo vortex, that is dominating

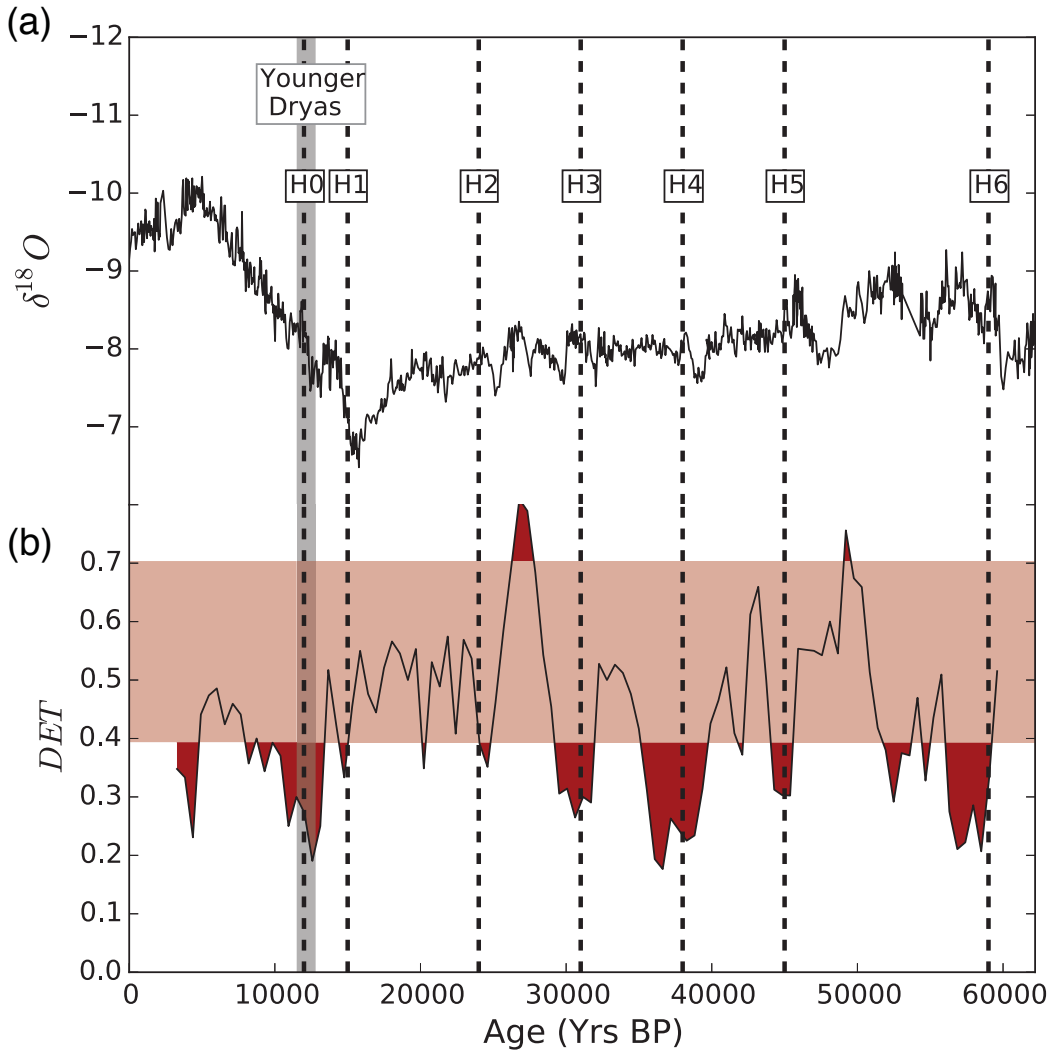

Figure 5.12: Upper graph: δ^{18}O record of the Secret Cave, Borneo; lower graph DET determined from the corresponding transformation costs time series. Horizontal lines H1 to H6 indicate the six Heinrich events while the most recent line determines the Younger Dryas, a cold period in the Northern Hemisphere most likely caused by a collapse of the North American ice sheet. The light red band of the DET indicates the 90% confidence interval.

the monsoon in more recent times (Koseki et al., 2014).

5.4.2 The See-Saw Relationship Between The Holocene East Asian–Australian Summer Monsoon

The East Asian–Indonesian–Australian monsoon (EAIAM) regime is the largest of the low–latitude monsoon systems. It links the Northern and Southern Hemispheres,

providing a planetary-scale heat source that drives the global circulation during boreal winter (McBride, 1987; Chang et al., 2006). As an important control on the dynamics of the monsoon system, at seasonal and inter-seasonal time scales, much has been made of the relationships between its geographical end members - the Indonesian–Australian summer monsoon (IAM) and East Asian summer monsoon (EAM) regimes (Suppiah, 1992; Chang et al., 2004; Wheeler and Mcbride, 2011). On these short-time scales, the summer monsoon hemisphere operating within the general seasonal progression of the monsoon (Chang et al., 2006) is linked via outflows from the winter monsoon of the opposing hemisphere (Suppiah and Wu, 1998; Chang et al., 2006). While such short time scale relationships have received considerable attention, likely phase relationships between the EASM and IASM over longer time scales are only starting to be deciphered (An, 2000; Wyrwoll et al., 2007; Ayliffe et al., 2013; Denniston et al., 2013; Mohtadi et al., 2014). With this uncertainty come questions of the likely long term adjustments to future greenhouse-triggered climate change, and whether these changes can 'lock-into' possible long term phase relationships between the Indonesian- Australian monsoon (IAM) and East Asian monsoon (EAM) regimes. Establishing such relationships would provide a significant step in understanding the climate system for a region where the lives of billions depend on monsoon–related rainfall.

The results of this work provide a step towards a better understanding of the centennial– to millennial–scale relationships within the EAIAM regime. I achieve this by using the TACTS method (Ozken et al., 2015) which enable us to: (i) confidently identify monsoon regime changes at the millennial to sub-centennial time scales in East Asian and Australian summer monsoon proxy records; (ii) demonstrate that the relationship over some 9000 years has taken on a see–saw relationship – with monsoon states essentially opposingly phased; and (iii) attribute this inter-hemispheric linkage to the solar variability that is impacting on both monsoon systems.

KNI-51 is located at the northern limits of Western Australia. Nearby Carlton Hill (15.49°S, 128.53°E) has a precipitation record extending back to 1897, with a mean annual rainfall over this period of 830mm (highest 1500mm/lowest 378mm) with an average 690 mm (83%) received during the monsoon season (December – March) (*Bureau of Meteorology,* 2013) . Mean annual precipitation near Dongge Cave is 1753 mm with 80% of the rainfall falling during the monsoon season – May to October (Dykoski et al., 2005). The high resolution speleothem paleoproxy records of KNI-51 and Dongge Cave outline the summer monsoon states of the last c. 9000 years.The details of the U/Th chronology and associated stable isotope records are provided by Denniston et al. (Denniston et al., 2013) and Wang et al. (Wang et al., 2005) respectively. Both locations are well placed to capture the respective summer monsoon regimes located at the end points of the broader EAIAM system (Fig. 5.13).

The records of Dongge Cave and KNI-51, as with many paleoclimate proxy records, are irregularly sampled, i.e.: the time between two consecutive measurements is not constant and may vary largely along the length of the record. Therefore, in order to preprocess this records, I applied the TACTS. For each proxy record, the detrended time series is divided into segment sizes of 20 years containing, on average, 4 to 5

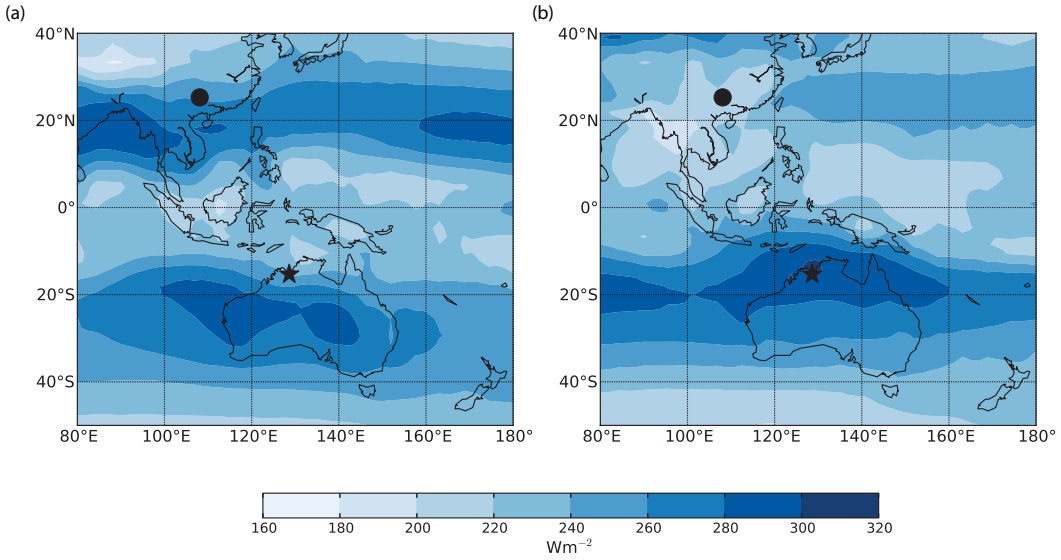

Figure 5.13: Top of atmosphere outgoing long wave radiation and 850mb vector winds delimiting the extent of: a) East Asian summer monsoon (JJA); and b) Indonesian Australian summer monsoon (DJF). Dongge Cave (dot) and KNI-51 (star)

points. The final results are relatively insensitive to the choice of segment size. The proportionality parameters for modifications (i) and (ii) are determined from the proxy records and are related to the average amplitude and sampling time. The creation and deletion cost factor is our optimisation parameter, chosen relative to the other parameters. Determining the costs of transformation provides a measure of how close one segment is to the following one and produces a regularly sampled transformation cost time series with a temporal resolution of 20 years. I analyze the time series using recurrence plot analysis and derive the determinism (Marwan et al., 2007), a measure related to predictability and well suited for detecting regime changes.

The wet/dry regimes identified by TACTS method (Fig. 5.14) can improve upon previous, qualitative interpretations of the proxy records Denniston et al., 2013; Wang et al., 2005; Hu et al., 2008. Here I provide a detailed discussion where our method supports, corrects and improves the analysis and interpretation of previous studies. I particularly focus on regimes which are newly identified or previously wrongly interpreted.

In the KNI-51 record (northwest Australian summer monsoon) the major wet (dry) phases occur between 8.5-6.4 ka (6.3-5.0 ka) 5.0-4.0 ka, possibly to 3.0 ka, (3.0-1.4 ka), 1.3-0.9 ka, with and from 0.9 ka a transition to the present monsoon regime. Embedded within these time intervals are additional events of centennial to sub–centennial duration. The major phase differences of the analysis show some correspondence with inferences drawn from a Holocene pollen/sediment record of monsoon events (McGowan et al., 2012), but with our analyses providing much

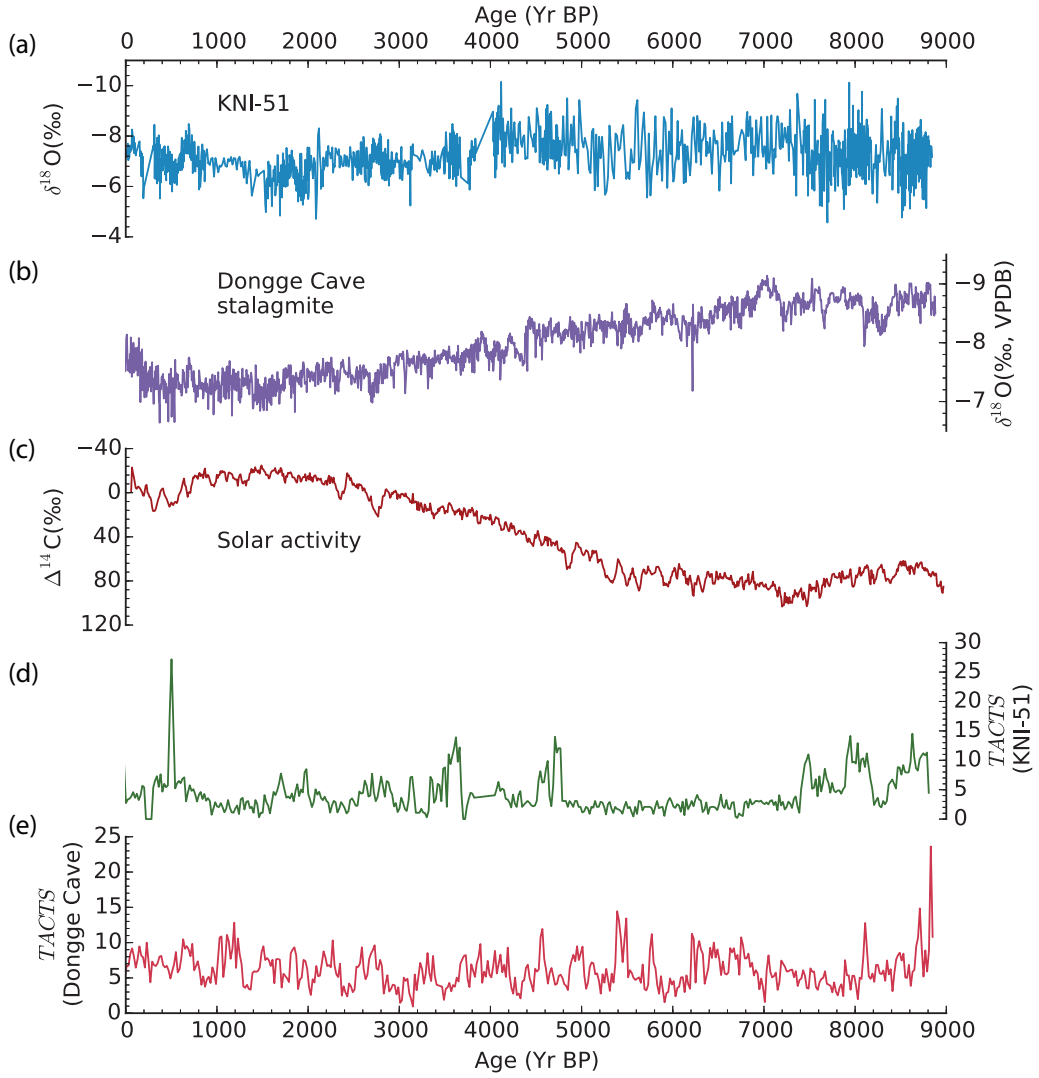

Figure 5.14: δ^{18} O records of (a) KNI-51 and (b) Dongge Cave. (c) atmospheric $\Delta^{14}C$ record and the transformation cost time series of (d) KNI-51 and (e) Dongge Cave determined by application of Eq. (5.6) and (5.7) to the detrended δ^{18} O records.

improved time resolution and a better resolution of details of the inherent variability within major monsoon phases.

The analysis of the Dongge Cave and KNI-51 records show statistically significant strong/weak monsoon states of centennial to millennial durations (Fig.5.15). The shaded bands in the figure depict 90% confidence intervals (calculated by bootstrapping method as in the previous application (Marwan et al., 2013)), with wet/dry states defined as exceeding these bands.

The Dongge Cave monsoon record has been discussed in detail by (Wang et al., 2005) and further developed by (Hu et al., 2008). Wang *et al.* recognized eight weak monsoon events lasting 100 to 500 years: at 0.5 ka, 1.6 ka, 2.7 ka, 4.4 ka, 5.5 ka,, 6.3 ka, 7.2 ka and 8.3 ka. While adding some details, Hu *et al.* reconstructions essentially concur with those of Wang et al. Our results indicate the wet (dry) regime intervals between (8.2-7.6 ka), 7.6-7.2 ka, (7.1-6.9 ka), (6.4-5.8 ka), 5.8-5.0 ka, (5.0-4.0 ka), 3.0-2.7 ka, (2.2-2.0 ka), 1.9-0.8 ka and (0.7-0.4 ka).

This analysis has revealed details for KNI-51 and Dongge Cave not previously recognized (Fig.5.15). In the KNI-51 record two events, absent from Denniston *et al.*, occur at 6.4-6.6 ka (wet) and 6.8-7.0 ka (dry). The results do not support the findings of Denniston *et al.* and McGowan *et al.* for the periods 3.1-3.2 ka (wet) and 7.5-7.6 ka (dry). Similarly, the results of our Dongge Cave analysis contradict the conclusions of Hu *et al.* for the time periods 6.1-6.2 ka (dry) and 7.6-7.8 ka (dry). In addition there are three events identified by Hu *et al.* that are not statistically significant in our analysis (3.2-3.4 ka, 6.3-6.9 ka and 8.2-8.8 ka). I assert confidence in these revisions, as they are based on a rigorous, quantitative analysis, rather than rudimentary visual comparison of data-sets.

The results reveal a striking wet–dry, opposing relationship between the IASM (Denniston et al., 2013) and EASM (Wang et al., 2005) (Fig. 5.15). The only time when this see–saw relationship is not observed is during 7.2-7.6 ka, when both monsoon records show a 'wet state'. Over the entire time scale, the cross–correlation of the determinism time series is -0.27, and while this affirms an antiphased relationship, it does not capture the strong correspondence between the statistically significant wet/dry monsoon states. In fact the antiphased relationship is much stronger, if only the statistical significant parts of the time series are used and the internal variability on sub–centennial to decadal time scales is ignored. While such an analysis can be easily done using a step function filter, I present here only the results without further filtering. Therefore the variability at sub–centennial to decadal time scales in both the Dongge Cave and KNI-51 records is emphasised; such short–term variability is evident in present day monsoon records from both regions (Webster, 2006).

While the details of the controls and processes determining the function and latitudinal extent of the respective summer monsoons are more complex (McBride, 1987; Chang et al., 2006; Chang et al., 2004) than simply relating them to the position of the Intertropical Convergence Zone (ITCZ), the ITCZ provides a convenient metric of monsoon extent (McBride, 1987; Waliser and Gautier, 1993).

The argument recognises that the ITCZ is displaced towards the warmer hemisphere in response to differential cooling (Chiang, 2003; Broccoli et al., 2006; Donohoe et al.,

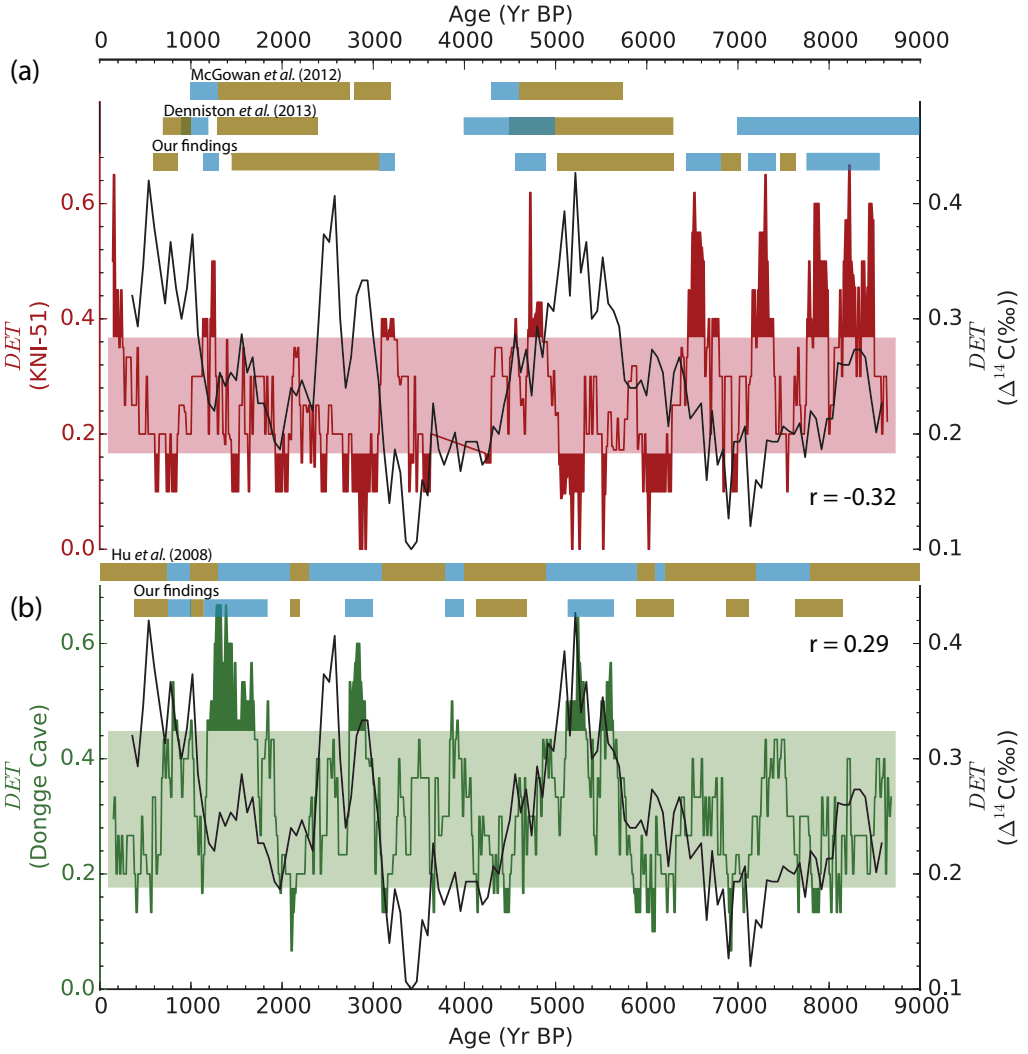

Figure 5.15: Determinism of the two proxy records (a) (red) KNI-51 and (b) (green) Dongge Cave. The determinism is calculated from the corresponding transformation costs time series and statistical significance is indicated by the two horizontal bands. High (low) determinism values correspond to wet (dry) monsoon regime. The coloured bands (blue indicating wet regimes; brown, dry) provide a comparison of our findings with those of previous qualitative studies. (black) Determinism of the solar activity proxy Δ^{14} C time series. Cross–correlation between the respective determinism of the solar activity proxy Δ^{14} C time series and KNI-51 time series is $r = -0.32$, and Dongge Cave time series is $r = 0.29$.

2013). This is an attractive and apparently straightforward explanation, with a caveat that the ITCZ over the region of the West Pacific Warm Pool (i.e. the Maritime Continent) is much less well defined than over the wider Pacific and Indian Oceans, with a more complex south–north (north–south) seasonal migration pattern (Waliser and Gautier, 1993; Hung and Yanai, 2004; Xian and Miller, 2008).

In explaining the Dongge Cave δ^{18}O record, Wang *et al.* appeal to a likely displacement of the ITCZ driven by solar variability. Their basis for this claim is the use of the atmospheric Δ^{14}C record as a proxy for solar activity, with which they obtain a correlation of 0.3 with their speleothem δ^{18}O record. We extend this claim further and ask whether the Holocene antiphase relationship that we have uncovered in the summer monsoons of the overall EAIAM is driven by solar variability.

To compare the KNI-51 and Dongge Cave records with solar variability (see Fig. 5.14 (a,b)), I correlate these records with the atmospheric $\Delta^{14}C$ record compiled by (Stuiver et al., 2006) (see Fig. 5.14 (c)). This record, spanning 9,700 years, was compiled from radiocarbon tree ring ages and is a widely used proxy for solar irradiance with lower $\Delta^{14}C$ values inferring increased solar irradiance (Wang et al., 2005). This record is already sampled at regular time intervals so I do not need to apply the transformation cost function. However, the time steps of this data set do not align with the determinism time series generated from the speleothem records. I cannot, therefore, calculate cross correlation without transforming the data sets again.

Interpolation is commonly used in such a scenario, but this creates artificial, and necessarily erroneous, data points into the time series. A kernel-based technique is therefore preferred, and the Gaussian kernel based cross correlation (gXCF) has been demonstrated to outperform linear interpolation, as well as Lomb-Scargle, rectangular and quasi-sinusoidal kernel based cross correlation estimators (Rehfeld et al., 2011). I therefore use gXCF as the estimator of the similarity between the speleothem and solar activity data sets.

The benefit of kernel based techniques is that, rather than introducing new data to the time series, the two data sets are 'matched' using a weighting function. As it is known well, traditional cross correlation takes the sum of the product of paired data points in two time series X and Y. However, here each data point in time series X is multiplied by *every* data point in Y, but with a weighting function dependent on the distance between the time that these observations occurred. Kernel based cross correlation is therefore given by

$$\hat{\rho}_{x,y} = \frac{\sum_i^{N_x} \sum_j^{N_y} x_i y_j b(t_j^y - t_i^x)}{\sum_i^{N_x} \sum_j^{N_y} b(t_j^y - t_i^x)} \tag{5.10}$$

where $b(t_j^y - t_i^x)$ is the kernel, determining how much weight to give to the product of two observations x_i and y_j, based on the time gap between them.

In the case of gXCF, the kernel is

$$b = \frac{1}{\sqrt{2\pi}\sigma} e^{-|d|^2/2\sigma^2} \tag{5.11}$$

where d is the distance between the observation times Δt_{ij}^{xy} and σ is the standard deviation of the kernel distribution, which scales the kernel. As there is no theory detailing the best choice of scaling parameter σ, I use $\sigma = \Delta t^{xy}/4$ as per (Rehfeld et al., 2011).

The analysis shows a statistically significant correlation between solar activity and both records from Dongge Cave (0.29) and KNI-51 (-0.32). Thus, when predictability of solar activity is high (low) and the Dongge Cave record indicates a strong (weak) summer monsoon, while northern Australia experiences a weak (strong) summer monsoon. These findings lead us to conclude that solar activity provides a driver in the see–saw relationship observed between the EASM and IASM over the past 9000 years.

A significant body of work is now available that proposes the impact of solar variability on the tropical atmosphere (Loon et al., 2004; Shindell et al., 2006; Lee et al., 2009; Meehl et al., 2009). This work demonstrates that the Hadley and Walker circulation cells are affected by solar variability, and through this, triggering an increase in tropical precipitation during solar maxima and bringing about an associated change in the position of the ITCZ. I demonstrate that solar variability can impact on summer monsoon strength, and more importantly provides the control of the antiphase relationship between the EASM and IASM over the last 9000 years. The results reveal a strong coupling between the monsoons of the two hemispheres – expressed as a see–saw relationship – and driven by decadal– to centennial–scale variations in solar activity.

5.5 Summary

In this chapter I have presented a novel method for analyzing irregularly sampled time series. This transformation cost time series (TACTS) method is based on determining similarities in time series segments. The fundamental transformation of the segments follows several elementary steps like moving a data point in time or changing its amplitude. By analyzing the average sampling rate and the average amplitude one can determine the associated cost factors for these transformation steps. Moreover as I have demonstrated the deletion and creation cost can be optimized relative to the first two costs. The advantage of the method is that the resulting transformation cost time series is regularly sampled. Therefore one is free to use a suitable time series method for further analysis without the risk of data corruption arising from unsuitable interpolation methods.

The extensive tests of the method have demonstrated, that TACTS is useful even for extreme irregular sampling and in addition can deal with rather high measurement noise. It can be used in discrete and continuous systems and shows promising results

when applied to real world applications. In combination with recurrence plot analysis measures like the determinism DET our method can detect dynamical regime changes accurately. Especially in areas like palaeo-climate, where often irregular sampling and parameter changes are common, our method provides a quantitative and objective way to analyze data and can guide scientists to previous hidden regime changes.

In the second application to palaeo-climate proxies, I have presented an interesting relationship between summer monsoon activities. The East Asian–Indonesian–Australian monsoon is the predominant low latitude monsoon system, providing a major global scale heat source. Here I have applied the TACTS method on speleothem climate proxies, from eastern China and northwestern Australia and establish relationships between the two summer monsoon regimes over the last ∼9000 years. I identify significant variations in monsoonal activity, both dry and wet phases, at millennial to multi-centennial time scales and demonstrate for the first time the existence of a see–saw antiphase relationship between the two regional monsoon systems. The analysis show that solar variability provides a likely mechanism driving the phase relationships.

Chapter 6

Discussion and Conclusion

Discussion

This Thesis mainly presented improved techniques for detection of abrupt changes in dynamical systems. Paleoclimate dynamics has been considered as the main aim of this thesis. In detail, I discussed the importance of dynamical regime transitions in paleoclimate data. When working with paleoclimate proxies we come across some difficulties similar to the once encountered in fields like astrophysics and the more broader earth sciences. In this work, I studied to overcome such problems and how to detect transitions from these time series via recurrence based techniques. In order to verify the yield of the newly developed techniques, underlying dynamics well-known synthetical data sets are created. The methods are experienced on the synthetical time series and used to uncover the transitions in real-world applications.

In the following I will summarize this thesis with brief discussion of the results.

Methodological background

The methodological background for this thesis is in the chapter two. The history and the fundamental properties of dynamical systems, important definitions for characterization, and the key tool, called recurrence plot (RP), for the thesis are given in this chapter. RP is a binary image and measures of RP are statistical computations from the structures in RP. These structures are black dots, black diagonal and vertical lines. I have reviewed the most popular measures and showed that they are able to detect the transitions between chaos-periodic, periodic-chaos and chaotic-chaotic regimes with a prototypical model.

Entropy of recurrence based approaches

Entropy of recurrence plot is the milestone of this chapter. Therefore, firstly I have briefly introduced the meaning of entropy and the expected behaviour for entropy at certain regimes. In the literature, the Shannon entropy of the recurrence matrix has been defined as a complexity measure and compared with the Shannon entropy of other recurrence-based approaches. Although entropy is a well known measure of disorder, in recurrence plot terminology, entropy is determined as a heuristic measure, in order to detect transitions between different regimes. The probability of occurrence of diagonal line segments of different lengths is not equal, since a recurrence plot is

a square matrix whose dimension is limited by the length of the time series. The Shannon entropy is computed from the diagonal line distribution in the RP approach. Hence, the commonly adopted entropic measures based on line segments can often yield counterintuitive results when quantifying the complexity of a given system. This was exemplified with the logistic map case in which the entropy of black and white dots was observed to be anticorrelated with the Lyapunov exponent. On the other hand, density based entropy of RPs presented here recovered the expected dependence as a function of the system's complexity, i.e., showing higher values within regions in which chaos is observed.

The arbitrary distance threshold of RP is a handicap for the analysis, since for different threshold values different RP are obtained. So far there is no unique way to select this threshold value. In order to get rid of this handicap, I have introduced the weighted recurrence plot (wRP). The density based entropy of wRP is applied on the logistic map. The outcome showed that the wRP entropy is a powerful measure and well suited to detect transitions in the underlaying dynamics. Moreover, for continuous systems such as the Rössler attractor and experimental time series of electrochemical oscillators, although black dots and weighted entropies are both positively correlated with the emergence of chaotic behavior, the latter definition was observed to have more stable values for voltage ranges that lead to periodic time series. The ideas presented here can be extended and applied to other complex systems with the potential to better identify dynamical transitions in time series originating from them. Specifically, where the Lyapunov exponent calculation is not possible due to limited number of time series points, of real data sets, the entropy of wRP detects different regimes very significantly.

Threshold selection for recurrence networks

Inspired by the increasing popularity of complex networks theory, I have studied a recurrence based network approach to identify the transitions in data sets. A novel method to chose the recurrence threshold is introduced. Before the newly developed method, I have introduced the network theory and some properties of networks very briefly, since these properties are used in the new threshold selection method. The technique to determine such a threshold automatically is completely depending on the time series itself. The threshold is directly derived from the topology of the recurrence network (RN). The idea behind the new method is that one should not miss any information from the time series. Which means, one should use entire time series in the network structure. If there are any unconnected components in the network, some network measures can give wrong or meaningless outcomes. On the other hand, one can consider the giant component as the recurrence network of time series, while missing some information from other part of the time series. In order to get rid of these problems, I introduced a method which selects the threshold in such a way that the recurrence network does not have any unconnected components.

Transitivity and betweenness centrality measures of the complex network have been used in the analysis. The adaptive threshold selection and the arbitrarily selected

threshold have compared by applying them to the logistic map. Although both methods distinguish the dynamical regime transitions between chaos to periodic or periodic to chaos clearly, the adaptively chosen threshold approach detects much more bifurcations, in particular such as period doubling. Such bifurcations are important characteristics of dynamical systems, since these bifurcations route to chaos via period doubling.

Moreover, a palaeoclimate proxy record from the palaeolake Lisan is analyzed with the adaptively selected threshold method. The data represents the climate variability in the near East between 27 and 18 cal ka BP. Both transitivity and betweenness centrality measures clearly identified transitions between wet and dry (and vice versa) periods by an abrupt decrease of dynamical regularity, perhaps due to a reduced solar influence. The analysis detected some transitions which have not been known so far in the literature and require further investigation, e.g., by analyzing other proxy records from this region. By choosing the adaptive threshold, I have been able to identify the transitions more clearly than by using the arbitrary selected threshold approach.

The transformation-cost time-series (TACTS)

Comprehensively understanding of the past has always been a thrilling question for scientist. One of the most important information source for the past is paleoclimate speleothem records. There are many traditional time series analysis techniques to uncover the underlying information of regularly sampled time series. But analyzing paleoclimate proxies is not always straight forward since, in general, data sets come with sampling irregularities. In other words, the sampling rate of such proxies is not constant. Therefore, these records are not directly analyzable. In order to work on such time series, one needs to apply some preprocessing procedures. Optimal preprocessing is the one which deforms the time series with the lowest damage. However, regularly interpolation as a preprocessing step is often applied, but this might lead to a serious bias of the results. An interpolation replaces real measurements with new interpolated data points close by that have regular sampling. These interpolated values have a higher uncertainty than the measured data points they replace.

In this work, I have introduced a new technique to analyze irregularly sampled time series. This transformation cost time series (TACTS) method is based on determining similarities in time series segments. As a result of this newly developed method, I showed that TACTS is useful even for extreme irregular sampling and in addition can deal with rather high measurement noise. It can be used in discrete and continuous systems and shows promising results when applied to real world applications. Therefore, one can perform an analysis on irregularly sampled time series without the risk of data corruption arising from interpolation.

The TACTS is a preprocessing method and after applying it, one can use any traditional analyzing procedure. In this work, I used RP measures to detect regime changes and found very accurate results. Especially in areas like palaeoclimate, where often irregular sampling and parameter changes are common, our method provides a

quantitative and objective way to analyze data and can guide scientists to previous hidden regime changes.

Via using combination of the TACTS method and determinism measure of RP, I have presented a very interesting relationship between summer monsoon activities. Firstly, the TACTS method was applied on two speleothem climate proxies, from eastern China and northwestern Australia and establish relationships between the two summer monsoon regimes over the last ∼9000 years. After having regularly sampled data sets, the variation of monsoon activity, both dry and wet regimes, has detected with the RP. For the first time, I have demonstrated the existence of a see–saw antiphase relationship between the two regional monsoon systems. Moreover, our analysis showed that solar variability provides a likely mechanism driving the phase relationships.

Conclusion

Although traditional recurrence based time series analysis has been demonstrated itself to be relevant in the detection of dynamical transitions, modifications on the methods are always necessary depending on each specific time series. Along with the improvements on recurrence based methods this work also suggests new measures, a new tool, new preprocessing and optimization techniques. The systematic comparison of the effects of these techniques on pragmatic models and real-world applications has shown that the results are much accurate and suitable than the traditional ways. Strikingly, using these methods we identified the see-saw relationship between the holocene East Asian–Indonesian–Australian summer monsoon (EAIAM). Moreover, further analysis for finding the reason of this relationship showed that the summer monsoons of the overall EAIAM is driven by solar variability.

Dynamics of climatic systems are very complex for instance a specific location is in relation with many different regions. Although I have studied just a relationship between two proxies, there are many interactions towards these proxies from different sources. In future work, we will consider many more proxies and try to find a network of relationships. This work can shed light on the reason of annual variations in climate change due to the effect of different links.

Open theoretical questions include selecting optimal parameters such as the total number of events in the windowing processes, the size of sliding windows, the size of their overlap, etc. especially for irregularly sampled time series analysis. A forthcoming theoretical study will try to optimize such arbitrary parameters regarding to a specific time series.

Although these techniques are used to distinguish the transitions in dynamical systems, their applications are not limited with just the detection of dynamical changes. They can be used in large variety of time series analysis methods. Analyzing a time series with its distinctive features might be better than getting trapped in the shortcomings of traditional methods.

Bibliography

Acheson, DJ (1990). *Elementary fluid dynamics: Oxford applied mathematics and computing science series*.

Afraimovich, V. (1997). "Pesin's dimension for Poincaré recurrences". In: *Chaos* 7.1, pp. 12–20. DOI: 10.1063/1.166237.

Afraimovich, V., J.-R. Chazottes, and B. Saussol (2000). "Local dimensions for Poincaré recurrences". In: *Electronic Research Announcements of the American Mathematical Society* 6, pp. 64–74.

– (2003). "Pointwise dimensions for Poincaré recurrences associated with maps and special flows". In: *Discrete and Continuous Dynamical Systems A* 9.2, pp. 263–280.

An, Zhisheng (2000). "The history and variability of the East Asian paleomonsoon climate". In: *Quaternary Science Reviews* 19, pp. 171–187. ISSN: 02773791. DOI: 10.1016/S0277-3791(99)00060-8.

Ayliffe, Linda K, Michael K Gagan, Jian-xin Zhao, Russell N Drysdale, John C Hellstrom, Wahyoe S Hantoro, Michael L Griffiths, Heather Scott-Gagan, Emma St Pierre, Joan a Cowley, and Bambang W Suwargadi (2013). "Rapid interhemispheric climate links via the Australasian monsoon during the last deglaciation." In: *Nature Communications* 4.May, p. 2908. ISSN: 2041-1723. DOI: 10.1038/ncomms3908.

Baptista, M. S., E. J. Ngamga, P. R. F. Pinto, M. Brito, and J. Kurths (2010). "Kolmogorov-Sinai entropy from recurrence times". In: *Physics Letters A* 374.9, pp. 1135–1140. DOI: 10.1016/j.physleta.2009.12.057.

Barreira, L. and B. Saussol (2001). "Hausdorff Dimension of Measures via Poincaré Recurrence". In: *Communications in Mathematical Physics* 219.2, pp. 443–463. DOI: 10.1007/s002200100427.

Barreira, Luis (2006). "Poincaré recurrence: old and new". In: *XIVth International Congress on Mathematical Physics*. Chap. 39, pp. 415–422. DOI: 10.1142/9789812704016_0039.

Bartov, Y, A Agnon, Y Enzel, and M Stein (2003). "Catastrophic arid episodes in the Eastern Mediterranean climate linked with the North Atlantic Heinrich events". In: *Geology* 31.5, pp. 439–442.

Berger, W.H. (1992). "The Younger Dryas cold spell - a quest for causes". In: *Nature* 360, pp. 219–237.

Boccaletti, S, V Latora, Y Moreno, M Chavez, and D U Hwang (2006). "Complex networks: structure and dynamics". In: *Physics Reports* 424, pp. 175–308. DOI: 10.1016/j.physrep.2005.10.009.

Bond, G., H. Heinrich, W. Broecker, L. Labeyrie, J. Mcmanus, J. Andrews, S. Huon, R. Jantschik, S. Clasen, and C. Simet (1992). "Evidence for massive discharges of

icebergs into the North Atlantic ocean during the last glacial period". In: *Global and Planetary Change* 360, pp. 245–249.

Bond, Gerard, William Showers, Maziet Cheseby, Rusty Lotti, Peter Almasi, Peter deMenocal, Paul Priore, Heidi Cullen, Irka Hajdas, and Georges Bonani (1997). "A Pervasive Millennial-Scale Cycle in North Atlantic Holocene and Glacial Climates". In: *Science* 278.5341, pp. 1257–1266. DOI: 10.1126/science.278.5341.1257. eprint: http://www.sciencemag.org/content/278/5341/1257.full.pdf.

Broccoli, Anthony J., Kristina a. Dahl, and Ronald J. Stouffer (2006). "Response of the ITCZ to Northern Hemisphere cooling". In: *Geophysical Research Letters* 33, pp. 1–4. ISSN: 00948276. DOI: 10.1029/2005GL024546.

Bureau of Meteorology, (2013). Climate Data Online, http://www.bom.gov.au/climate/data/.

Carolin, Stacy A., Kim M. Cobb, Jess F. Adkins, Brian Clark, Jessica L. Conroy, Syria Lejau, Jenny Malang, and Andrew A. Tuen (2013). "Varied Response of Western Pacific Hydrology to Climate Forcings over the Last Glacial Period". In: *Science* 340, pp. 1564–1566.

Chang, C.-P., P.A. Harr, J. McBride, and H.-H. Hsu (2004). "Maritime continent monsoon; annual cycle and boreal winter variability". In: *East Asian Monsoon. World Scientific Series on Meteorology of East Asia, vol. 2*. Ed. by C.-P. Chang. Singapore: World Scientific Publishing Co. Pte. Ltd, pp. 107–150.

Chang, C.-P., Zhuo Wang, and Harry Hendon (2006). "The Asian winter monsoon". English. In: *The Asian Monsoon*. Springer Praxis Books. Springer Berlin Heidelberg, pp. 89–127. ISBN: 978-3-540-40610-5. DOI: 10.1007/3-540-37722-0_3.

Chialvo, D R, R F Gilmour, and J Jalife (1990). "Low dimensional chaos in cardiac tissue." In: *Nature* 343.6259, pp. 653–657. ISSN: 0028-0836. DOI: 10.1038/343653a0.

Chiang, John C. H. (2003). "Sensitivity of the Atlantic Intertropical Convergence Zone to Last Glacial Maximum boundary conditions". In: *Paleoceanography* 18.4, pp. 1–18. ISSN: 0883-8305. DOI: 10.1029/2003PA000916.

Denniston, Rhawn F., Karl Heinz Wyrwoll, Victor J. Polyak, Josephine R. Brown, Yemane Asmerom, Alan D. Wanamaker, Zachary LaPointe, Rebecca Ellerbroek, Michael Barthelmes, Daniel Cleary, John Cugley, David Woods, and William F. Humphreys (2013). "A Stalagmite record of holocene indonesian-australian summer monsoon variability from the australian tropics". In: *Quaternary Science Reviews* 78, pp. 155–168. ISSN: 02773791. DOI: 10.1016/j.quascirev.2013.08.004.

Donges, J. F., J. Heitzig, R. V. Donner, and J. Kurths (2012). "Analytical framework for recurrence network analysis of time series". In: *Physical Review E* 85, p. 046105. DOI: 10.1103/PhysRevE.85.046105.

Donner, R. V., Y. Zou, J. F. Donges, N. Marwan, and J. Kurths (2010). "Ambiguities in recurrence-based complex network representations of time series". In: *Physical Review E* 81, 015101(R). DOI: 10.1103/PhysRevE.81.015101.

Donner, Reik V. and Jonathan F. Donges (2012). *Visibility graph analysis of geophysical time series: Potentials and possible pitfalls*. Vol. 60. 3, pp. 589–623. ISBN: 1160001200. DOI: 10.2478/s11600-012-0032-x.

Donohoe, Aaron, John Marshall, David Ferreira, and David McGee (2013). "The relationship between ITCZ location and cross-equatorial atmospheric heat transport:

From the seasonal cycle to the last glacial maximum". In: *Journal of Climate* 26.2008, pp. 3597–3618. ISSN: 08948755. DOI: 10.1175/JCLI-D-12-00467.1.

Dykoski, Carolyn a., R. Lawrence Edwards, Hai Cheng, Daoxian Yuan, Yanjun Cai, Meiliang Zhang, Yushi Lin, Jiaming Qing, Zhisheng An, and Justin Revenaugh (2005). "A high-resolution, absolute-dated Holocene and deglacial Asian monsoon record from Dongge Cave, China". In: *Earth and Planetary Science Letters* 233, pp. 71–86. ISSN: 0012821X. DOI: 10.1016/j.epsl.2005.01.036.

Eckmann, J.-P, S. Oliffson Kamphorst, and D Ruelle (1987). "Recurrence Plots of Dynamical Systems". In: *Europhysics Letters (EPL)* 4.9, pp. 973–977. ISSN: 0295-5075. DOI: 10.1209/0295-5075/4/9/004.

Eroglu, D., N. Marwan, S. Prasad, and J. Kurths (2014a). "Finding recurrence networks' threshold adaptively for a specific time series". In: *Nonlinear Processes in Geophysics* 21.6, pp. 1085–1092. DOI: 10.5194/npg-21-1085-2014.

Eroglu, Deniz, Thomas K. DM. Peron, Nobert Marwan, Francisco A. Rodrigues, Luciano da F. Costa, Michael Sebek, István Z. Kiss, and Jürgen Kurths (2014b). "Entropy of weighted recurrence plots". In: *Phys. Rev. E* 90 (4), p. 042919. DOI: 10.1103/PhysRevE.90.042919.

Eroglu, Deniz, Fiona McRobie, Ibrahim Ozken, Thomas Stemler, Karl-Heinz Wyrwoll, Norbert Marwan, and Jurgen Kurths (2015). "The Holocene East Asian-Australian summer monsoon: A see-saw relationship".

Floriani, Leila De, Paola Marzano, and Enrico Puppo (1994). "Line-of-sight communication on terrain models". In: *International Journal of Geographical Information Systems* 8.4, pp. 329–342. DOI: 10.1080/02693799408902004.

Gao, J. B. (1999). "Recurrence Time Statistics for Chaotic Systems and Their Applications". In: *Physical Review Letters* 83.16, pp. 3178–3181. DOI: 10.1103/PhysRevLett.83.3178.

Grassberger, P. (1983a). "Generalized Dimensions of Strange Attractors". In: *Physics Letters A* 97.6, pp. 227–230. DOI: 10.1016/0375-9601(83)90753-3.

Grassberger, P. and I. Procaccia (1983a). "Characterization of strange attractors". In: *Physical Review Letters* 50.5, pp. 346–349. DOI: 10.1103/PhysRevLett.50.346.

– (1983b). "Estimation of the Kolmogorov entropy from a chaotic signal". In: *Physical Review A* 9.1–2, pp. 2591–2593. DOI: 10.1103/PhysRevA.28.2591.

– (1983c). "Measuring the strangeness of strange attractors". In: *Physica D* 9.1–2, pp. 189–208. DOI: 10.1016/0167-2789(83)90298-1.

– (1984). "Dimensions and entropies of strange attractors from a fluctuating dynamics approach". In: *Physics Letters A* 13.1–2, pp. 34–54. DOI: 10.1016/0167-2789(84)90269-0.

Grassberger, Peter (1983b). "Generalized dimensions of strange attractors". In: *Physics Letters A* 97.6, pp. 227–230. ISSN: 03759601. DOI: 10.1016/0375-9601(83)90753-3.

Hadyn, N., J. Luevano, G. Mantica, and S. Vaienti (2002). "Multifractal Properties of Return Time Statistics". In: *Physical Review Letters* 88.22, p. 224502. DOI: 10.1103/PhysRevLett.88.224502.

Hensman, James, Neil D Lawrence, and Magnus Rattray (2013). "Hierarchical Bayesian modelling of gene expression time series across irregularly sampled replicates and clusters". In: *BMC bioinformatics* 14.1, p. 252.

Hirata, Masaki, BenoÃ®t Saussol, and Sandro Vaienti (1999). "Statistics of Return Times:Â¶A General Framework and New Applications". English. In: *Communications in Mathematical Physics* 206.1, pp. 33–55. ISSN: 0010-3616. DOI: 10.1007/s002200050697.

Hirata, Y., S. Horai, and K. Aihara (2008). "Reproduction of distance matrices from recurrence plots and its applications". In: *European Physical Journal – Special Topics* 164.1, pp. 13–22. DOI: 10.1140/epjst/e2008-00830-8.

Hirata, Yoshito and Kazuyuki Aihara (2009). "Representing spike trains using constant sampling intervals". In: *Journal of Neuroscience Methods* 183.2, pp. 277 –286. ISSN: 0165-0270. DOI: http://dx.doi.org/10.1016/j.jneumeth.2009.06.030.

Hu, Chaoyong, Gideon M. Henderson, Shucheng Xie Junhua Huang, Ying Sun, and Kathleen R. Johnson (2008). "Quantification of Holocene Asian monsoon rainfall from spatially separated cave records". In: *Earth and Planetary Science Letters* 266, pp. 221–232.

Huang, K. (1987). *Statistical mechanics*. Wiley. ISBN: 9780471815181.

Hung, Chih-Wen and Michio Yanai (2004). "Factors contributing to the onset of the Australian summer monsoon". In: *Quarterly Journal of the Royal Meteorological Society* 130, pp. 739–758. ISSN: 1477-870X. DOI: 10.1256/qj.02.191.

Kac, M. (1947). "On the notion of recurrence in discrete stochastic processes". In: *Bull. Amer. Math. Soc.* 53.10, pp. 1002–1010.

Kantz, H. and T. Schreiber (1997). *Nonlinear Time Series Analysis*. Cambridge: University Press.

Katok, A. and B. Hasselblatt (1995). *Introduction to the Modern Theory of Dynamical Systems*. Cambridge: Cambridge University Press.

Kennel, M. B., R. Brown, and H. D. I. Abarbanel (1992). "Determining embedding dimension for phase-space reconstruction using a geometrical construction". In: *Physical Review A* 45.6, pp. 3403–3411. DOI: 10.1103/PhysRevA.45.3403.

Kiss, I. Z., W. Wang, and J. L. Hudson (2000). "Complexity of globally coupled chaotic electrochemical oscillators". In: *Phys. Chem. Chem. Phys.* 2, pp. 3847–3854.

Kiss, Istvan Z., Yumei Zhai, and John L. Hudson (2002). "Emerging coherence in a population of chemical oscillators". In: *Science* 296.5573, pp. 1676–1678.

Koebbe, M. and G. Mayer-Kress (1992). "Use of Recurrence Plots in the Analysis of Time-Series Data". In: *Proceedings of SFI Studies in the Science of Complexity*. Ed. by M. Casdagli and S. Eubank. Vol. XXI. Redwood City: Addison-Wesley, pp. 361–378.

Koper, M. T. M. (1996). "Oscillations and complex dynamical bifurcations in electrochemical systems". In: *Adv. Chem. Phys.* 92, pp. 161–298.

Koseki, S., T.Y. Koh, and C.K. Teo (2014). "Borneo vortex and mesoscale convective rainfall". In: *Atmospheric Chemistry and Physics* 14, pp. 4539–4562.

Kreindler, David M and Charles J Lumsden (2012). "The effects of the irregular sample and missing data in time series analysis". In: *Nonlinear Dynamical Systems Analysis for the Behavioral Sciences Using Real Data*, p. 135.

Lee, Jae N., Drew T. Shindell, and Sultan Hameed (2009). "The influence of solar forcing on tropical circulation". In: *Journal of Climate* 22.Berger 1978, pp. 5870–5885. ISSN: 08948755. DOI: 10.1175/2009JCLI2670.1.

Letellier, C. (2006). "Estimating the Shannon Entropy: Recurrence Plots versus Symbolic Dynamics". In: *Physical Review Letters* 96, p. 254102. DOI: 10.1103/PhysRevLett.96.254102.

Lev, O., A. Wolffberg, L. M. Pismen, and M. Sheintuch (1989). "The structure of complex behavior in anodic nickel dissolution". In: *J. Phys. Chem.* 93.4, pp. 1661–1666.

Liénard, A (1928). "Etude des oscillations entretenues". In: *Revue générale de l'électricité* 23.21, pp. 901–912.

Linsay, Paul S. (1981). "Period doubling and chaotic behavior in a driven anharmonic oscillator". In: *Physical Review Letters* 47.19, pp. 1349–1352. ISSN: 00319007. DOI: 10.1103/PhysRevLett.47.1349.

Little, L. M., P. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz (2007). "Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection". In: *BioMedical Engineering OnLine* 6.23, pp. 1–19. DOI: 10.1186/1475-925X-6-23.

Livina, V. N. and T. M. Lenton (2007). "A modified method for detecting incipient bifurcations in a dynamical system". In: *Geophysical Research Letters* 34.3, p. L03712. ISSN: 1944-8007. DOI: 10.1029/2006GL028672.

Loon, Harry van, Gerald A. Meehl, and Julie M. Arblaster (2004). "A decadal solar effect in the tropics in July-August". In: *Journal of Atmospheric and Solar-Terrestrial Physics* 66, pp. 1767–1778. ISSN: 13646826. DOI: 10.1016/j.jastp.2004.06.003.

Lozano-Pérez, Tomás and Michael A. Wesley (1979). "An Algorithm for Planning Collision-free Paths Among Polyhedral Obstacles". In: *Commun. ACM* 22.10, pp. 560–570. ISSN: 0001-0782. DOI: 10.1145/359156.359164.

Malik, Nishant, Norbert Marwan, Yong Zou, Peter J. Mucha, and Jürgen Kurths (2014). "Fluctuation of similarity to detect transitions between distinct dynamical regimes in short time series". In: *Phys. Rev. E* 89 (6), p. 062908. DOI: 10.1103/PhysRevE.89.062908.

Marwan, N. (2008). "A Historical Review of Recurrence Plots". In: *European Physical Journal – Special Topics* 164.1, pp. 3–12. DOI: 10.1140/epjst/e2008-00829-1.

Marwan, N., M. C. Romano, M. Thiel, and J. Kurths (2007). "Recurrence Plots for the Analysis of Complex Systems". In: *Physics Reports* 438.5–6, pp. 237–329. DOI: 10.1016/j.physrep.2006.11.001.

Marwan, N., J. F. Donges, Y. Zou, R. V. Donner, and J. Kurths (2009). "Complex network approach for recurrence analysis of time series". In: *Physics Letters A* 373.46, pp. 4246–4254. DOI: 10.1016/j.physleta.2009.09.042.

Marwan, N., S. Schinkel, and J. Kurths (2013). "Recurrence plots 25 years later – Gaining confidence in dynamical transitions". In: *Europhysics Letters* 101, p. 20007. DOI: 10.1209/0295-5075/101/20007.

Marwan, Norbert, Niels Wessel, Udo Meyerfeldt, Alexander Schirdewan, and Jürgen Kurths (2002). "Recurrence-plot-based measures of complexity and their application to heart-rate-variability data". In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 66, pp. 1–8. ISSN: 15393755. DOI: 10.1103/PhysRevE.66.026702.

McBride, John L. (1987). "The Australian summer monsoon". In: *Monsoon Meteorology*. Ed. by C. P. Chang and T. N. Krishnamurti. Oxford, U.K.: Oxford University Press, pp. 203–231.

McGowan, Hamish, Samuel Marx, Patrick Moss, and Andrew Hammond (2012). "Evidence of ENSO mega-drought triggered collapse of prehistory Aboriginal society in northwest Australia". In: *Geophysical Research Letters* 39.22, p. L22702. ISSN: 1944-8007. DOI: 10.1029/2012GL053916.

Meehl, Gerald A., Julie M Arblaster, Katja Matthes, Fabrizio Sassi, and Harry van Loon (2009). "Amplifying the Pacific climate system response to a small 11-year solar cycle forcing." In: *Science* 325.2009, pp. 1114–1118. ISSN: 0036-8075. DOI: 10.1126/science.1172872.

Michael T. Rosenstein, James J. Collins and Carlo J. De Luca (1993). "A practical method for calculating largest Lyapunov exponents from small data sets". In: *Physica D* 65.3, pp. 117–134. DOI: SDI:0167-2789(92)00033-6.

Mindlin, G. M. and R. Gilmore (1992). "Topological analysis and synthesis of chaotic time series". In: *Physica D* 58.1–4, pp. 229–242. DOI: 10.1016/0167-2789(92)90111-Y.

Mohtadi, Mahyar, Matthias Prange, Delia W Oppo, Ricardo De Pol-Holz, Ute Merkel, Xiao Zhang, Stephan Steinke, and Andreas Lückge (2014). "North Atlantic forcing of tropical Indian Ocean climate." In: *Nature* 509, pp. 76–80. ISSN: 1476-4687. DOI: 10.1038/nature13196.

Nagy, George (1994). "Terrain visibility". In: *Comp. and Grap.* 18.6, pp. 763 –773. ISSN: 0097-8493. DOI: http://dx.doi.org/10.1016/0097-8493(94)90002-7.

Newman, M. (2010). *Networks: An Introduction*. OUP Oxford. ISBN: 9780199206650.

Newman, M. E. J. (2002). "Assortative Mixing in Networks". In: *Phys. Rev. Lett.* 89 (20), p. 208701. DOI: 10.1103/PhysRevLett.89.208701.

– (2003). "Mixing patterns in networks". In: *Phys. Rev. E* 67 (2), p. 026126. DOI: 10.1103/PhysRevE.67.026126.

Oblow, E.M. (1988). "Supertracks, supertrack functions and chaos in the quadratic map". In: *Physics Letters A* 128.8, pp. 406 –412. ISSN: 0375-9601. DOI: http://dx.doi.org/10.1016/0375-9601(88)90119-3.

Ossadnik, S.M., S.V. Buldyrev, A.L. Goldberger, S. Havlin, R.N. Mantegna, C.K. Peng, M. Simons, and H.E. Stanley (1994). "Correlation approach to identify coding regions in DNA sequences". In: *Biophysical Journal* 67.1, pp. 64 –70. ISSN: 0006-3495. DOI: http://dx.doi.org/10.1016/S0006-3495(94)80455-2.

Ozken, Ibrahim, Deniz Eroglu, Thomas Stemler, Norbert Marwan, G. Baris Bagci, and Jürgen Kurths (2015). "Transformation-cost time-series method for analyzing irregularly sampled data". In: *Physical Review E* 91.6, p. 062911. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.91.062911.

Packard, N. H., J. P. Crutchfield, J. D. Farmer, and R. S. Shaw (1980). "Geometry from a Time Series". In: *Physical Review Letters* 45.9, pp. 712–716. DOI: 10.1103/PhysRevLett.45.712.

Pausata, Francesco S. R., David S. Battisti, Kerim H. Nisancioglu, and Cecilia M. Bitz (2011). "Chinese stalagmite δ18O controlled by changes in the Indian monsoon during a simulated Heinrich event". In: *Nature Geoscience* 4.7, pp. 474–480. ISSN: 1752-0894. DOI: 10.1038/ngeo1169.

Peng, C.-K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger (1994). "Mosaic organization of DNA nucleotides". In: *Phys. Rev. E* 49 (2), pp. 1685–1689. DOI: 10.1103/PhysRevE.49.1685.

Penné, V., B. Saussol, and S. Vaienti (1999). "Dimensions for recurrence times: topological and dynamical properties". In: *Discrete and Continuous Dynamical Systems A* 4.4, pp. 783–798.

Pincus, S M (1991). "Approximate entropy as a measure of system complexity". In: *Proceedings of the National Academy of Sciences of the United States of America* 88.6, pp. 2297–2301. ISSN: 0027-8424. DOI: 10.1073/pnas.88.6.2297.

Poincaré, H. (1890). "Sur la probleme des trois corps et les équations de la dynamique". In: *Acta Mathematica* 13, pp. 1–271.

Prasad, Sushma, Heinz Vos, J.F.W. Negendank, Nicholas Waldmann, Steven L. Goldstein, and Mordechai Stein (2004). "Evidence from Lake Lisan of solar influence on decadal- to centennial-scale climate variability during marine oxygen isotope stage 2". In: *Geology* 32.7, p. 581. ISSN: 0091-7613. DOI: 10.1130/G20553.1.

Prasad, Sushma, J.F.W. Negendank, and Mordechai Stein (2009). "Varve counting reveals high resolution radiocarbon reservoir age variations in palaeolake Lisan". In: *Journal of Quaternary Science* 24.7, pp. 690–696.

Rehfeld, K. and J. Kurths (2014). "Similarity estimators for irregular and age-uncertain time series". In: *Climate of the Past* 10.1, pp. 107–122. DOI: 10.5194/cp-10-107-2014.

Rehfeld, K., N. Marwan, S. F. M. Breitenbach, and J. Kurths (2012). "Late Holocene Asian summer monsoon dynamics from small but complex networks of paleoclimate data". In: *Climate Dynamics* 41.1, pp. 3–19. DOI: 10.1007/s00382-012-1448-3.

Rehfeld, Kira, Norbert Marwan, Jobst Heitzig, and Jürgen Kurths (2011). "Comparison of correlation analysis techniques for irregularly sampled time series". In: *Nonlinear Processes in Geophysics* 18.3, pp. 389–404. DOI: 10.5194/npg-18-389-2011.

Rodó, Xavier and Miquel-Angel Rodriguez-Arias (2006). "A new method to detect transitory signatures and local time/space variability structures in the climate system: the scale-dependent correlation analysis". English. In: *Climate Dynamics* 27.5, pp. 441–458. ISSN: 0930-7575. DOI: 10.1007/s00382-005-0106-4.

Rössler, O. E. (1976). "An equation for continuous chaos". In: *Physics Letters A* 57.5, pp. 397–398. DOI: 10.1016/0375-9601(76)90101-8.

Sakoe, Hiroaki and Seibi Chiba (1978). "Dynamic Programming Algorithm Optimization for Spoken Word Recognition". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26.1, pp. 43–49. DOI: 10.1007/s00442-003-1464-4.

Saussol, B. and J. Wu (2003). "Recurrence spectrum in smooth dynamical systems". In: *Nonlinearity* 16.6, pp. 1991–2001. DOI: 10.1088/0951-7715/16/6/306.

Saussol, B., S. Troubetzkoy, and S. Vaienti (2002). "Recurrence, Dimensions, and Lyapunov Exponents". In: *Journal of Statistical Physics* 106.3–4, pp. 623–634. DOI: 10.1023/A:1013710422755.

– (2003). "Recurrence and Lyapunov Exponents for diffeomorphisms". In: *Moscow Mathematical Journal* 3.1, pp. 189–203.

Schinkel, S., O. Dimigen, and N. Marwan (2008). "Selection of recurrence threshold for signal detection". In: *European Physical Journal – Special Topics* 164.1, pp. 45–53. DOI: 10.1140/epjst/e2008-00833-5.

Scott, J.F., M. Sargent, and C.D. Cantrell (1975). "Laser-phase transition analogy: Application to first-order transitions". In: *Optics Communications* 15.1, pp. 13 –16. ISSN: 0030-4018. DOI: http://dx.doi.org/10.1016/0030-4018(75)90171-6.

Shannon, C. E. (1948). "A Mathematical Theory of Communication". In: *Bell System Technical Journal* 27.3, pp. 379–423. ISSN: 1538-7305. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Shindell, Drew T., G Faluvegi, R.L. Miller, G.A. Schmidt, and J.E. Hansen (2006). "Solar and anthropogenic forcing of tropical hydrology". In: *Geophysical Research Letters* 33, p. L24706.

Sinai, Ya.G. (1959). "On the Notion of Entropy of a Dynamical System". In: *Doklady of Russian Academy of Sciences* 124, pp. 768–771.

Singer, W (1999). "Neuronal synchrony: a versatile code for the definition of relations?" In: *Neuron* 24.1, pp. 49–65, 111–125.

Sipers, A., P. Born, and R. Peeters (2011). "On the unique reconstruction of a signal from its unthresholded recurrence plot". In: *Physics Letters A* 375.24, pp. 2309–2321. DOI: 10.1016/j.physleta.2011.04.040.

Small, Michael, Dejin Yu, and Robert G. Harrison (2003). "Observation of a Period Doubling Bifurcation During Onset of Human Ventricular Fibrillation". In: *International Journal of Bifurcation and Chaos* 13.03, pp. 743–754. ISSN: 0218-1274. DOI: 10.1142/S0218127403006911.

Stein, M, A Starinsky, A Katz, SL Goldstein, M Machlus, and A Schramm (2003). "Strontium isotopic, chemical, and sedimentological evidence for the evolution of Lake Lisan and the Dead Sea". In: *Geochimica et Cosmochimica Acta* 61.18, pp. 3975–3992.

Stuiver, M and PM Grootes (2000). "GISP2 oxygen isotope ratios". In: *Quaternary Research* 53, pp. 277–283.

Stuiver, Minze, Paula Reimer, Edouard Bard, J Beck, G Burr, Konrad Hughen, Bernd Kromer, Gerry McCormac, Johannes van der Plicht, and Marco Spurk (2006). "INTCAL98 radiocarbon age calibration, 24,000-0 cal BP." In: *Radiocarbon* 40.3. ISSN: 0033-8222.

Suppiah, R. (1992). "The Australian summer a review Suppiah". In: *Progress in Physical Geography* 16.3, pp. 283–318.

Suppiah, Ramasamy and Xingren Wu (1998). "Surges, Cross-Equatorial Flows and Their Links with The Australian Summer Monsoon Circulation dan Rainfall". In: *Second Australian Conference on Agricultural Meteorology: conference proceedings: the impact of weather and climate on agriculture*, pp. 113–130.

Suzuki, S., Y. Hirata, and K. Aihara (2010). "Definition of distance for marked point process data and its application to recurrence plot-based analysis of exchange tick data of foreign currencies". In: *International Journal of Bifurcation and Chaos* 20.11, pp. 3699–3708. DOI: 10.1142/S0218127410027970.

Thiebaut, Carole and Sylvie Roques (2005). "Time-scale and time-frequency analyses of irregularly sampled astronomical time series". In: *EURASIP Journal on Applied Signal Processing* 2005, pp. 2486–2499.

Thiel, M., M. C. Romano, J. Kurths, R. Meucci, E. Allaria, and F. T. Arecchi (2002). "Influence of observational noise on the recurrence quantification analysis". In: *Physica D* 171.3, pp. 138–152. DOI: 10.1016/S0167-2789(02)00586-9.

Thiel, M., M. C. Romano, P. L. Read, and J. Kurths (2004a). "Estimation of dynamical invariants without embedding by recurrence plots". In: *Chaos* 14.2, pp. 234–243. DOI: 10.1063/1.1667633.

Thiel, M., M. C. Romano, and J. Kurths (2004b). "How much information is contained in a recurrence plot?" In: *Physics Letters A* 330.5, pp. 343–349. DOI: 10.1016/j.physleta.2004.07.050.

Torfstein, A, SL Goldstein, M Stein, and Y Enzel (2013). "Impacts of abrupt climate changes in the Levant from Last Glacial Dead Sea levels". In: *Quaternary Science Reviews* 69, pp. 1–7. ISSN: 0277-3791.

Trulla, L. L., A. Giuliani, J. P. Zbilut, and C. L. Webber Jr. (1996). "Recurrence quantification analysis of the logistic equation with transients". In: *Physics Letters A* 223.4, pp. 255–260. DOI: 10.1016/S0375-9601(96)00741-4.

Turner, Alasdair, Maria Doxa, David O'sullivan, and Alan Penn (2001). "From isovists to visibility graphs: a methodology for the analysis of architectural space". In: *Environ Plann B* 28.1, pp. 103–121.

Varga, Richard S (2010). *Geršgorin and his circles*. Vol. 36. Springer Science & Business Media.

Victor, J. D. and K. P. Purpura (1997). "Metric-space analysis of spike trains: theory, algorithms and application". In: *Network: Computation in Neural Systems* 8.2, pp. 127–164.

Waliser, D. E. and C. Gautier (1993). "A satellite-derived climatology of the ITCZ". In: *Journal of Climate* 6, pp. 2162–2174. ISSN: 08948755. DOI: 10.1175/1520-0442(1993)006<2162:ASDCOT>2.0.CO;2.

Wang, Yongjin, Hai Cheng, R. Lawrence Edwards, Yaoqi He, Xinggong Kong, Zhisheng An, Jiangying Wu, Megan J. Kelly, Carolyn A. Dykoski, and Xiangdong Li (2005). "The Holocene Asian Monsoon: Links to Solar Changes and North Atlantic Climate". In: *Science* 308.5723, pp. 854–857. DOI: 10.1126/science.1106296. eprint: http://www.sciencemag.org/content/308/5723/854.full.pdf.

Webber, C L and J P Zbilut (1994). "Dynamical assessment of physiological systems and states using recurrence plot strategies." In: *Journal of applied physiology (Bethesda, Md. : 1985)* 76.2, pp. 965–973. ISSN: 8750-7587.

Webster, Peter J. (2006). "The Asian Monsoon". English. In: *The Asian Monsoon*. Ed. by B. Wang. Springer Praxis Publishing, Chichester,UK, pp. 3–66.

Wheeler, M C and J L Mcbride (2011). "Australasian monsoon". In: *Intraseasonal Variability in the Atmosphere-Ocean Climate System (2nd edition)*. Ed. by W.K.M. Lau and D.E. Waliser. Springer, pp. 147–198.

Wyrwoll, Karl Heinz, Zhengyu Liu, Guangshan Chen, John E. Kutzbach, and Xiaodong Liu (2007). "Sensitivity of the Australian summer monsoon to tilt and precession forcing". In: *Quaternary Science Reviews* 26, pp. 3043–3057. ISSN: 02773791. DOI: 10.1016/j.quascirev.2007.06.026.

Xian, Peng and Ron L. Miller (2008). "Abrupt Seasonal Migration of the ITCZ into the Summer Hemisphere". In: *Journal of the Atmospheric Sciences* 65, pp. 1878–1895. ISSN: 0022-4928. DOI: 10.1175/2007JAS2367.1.

Xu, Xiaoke, Jie Zhang, and Michael Small (2008). "Superfamily phenomena and motifs of networks induced from time series." In: *Proceedings of the National Academy of Sciences of the United States of America* 105.50, pp. 19601–19605. ISSN: 0027-8424. DOI: 10.1073/pnas.0806082105.

Ye, Jun, Hua Li, and John G. McInerney (1993). "Period-doubling route to chaos in a semiconductor laser with weak optical feedback". In: *Physical Review A* 47.3, pp. 2249–2252. ISSN: 10502947. DOI: 10.1103/PhysRevA.47.2249.

Zbilut, Joseph P. and Charles L. Webber (1992). "Embeddings and delays as derived from quantification of recurrence plots". In: *Physics Letters A* 171.3-4, pp. 199–203. ISSN: 03759601. DOI: 10.1016/0375-9601(92)90426-M.

Zhang, J. and M. Small (2006). "Complex network from pseudoperiodic time series: Topology versus dynamics". In: *Physical Review Letters* 96.23, pp. 3–6. ISSN: 00319007. DOI: 10.1103/PhysRevLett.96.238701.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Potsdam, den 3. Februar 2016 Deniz Eroglu